Л.Б. Ряшко

ОПТИМАЛЬНОЕ РЕКУРРЕНТНОЕ ОЦЕНИВАНИЕ

Екатеринбург • 1997

Министерство общего и профессионального образования
Российской Федерации
Уральский государственный университет
им. А. М. Горького

Л.Б. Ряшко

ОПТИМАЛЬНОЕ РЕКУРРЕНТНОЕ ОЦЕНИВАНИЕ

Учебное пособие

Екатеринбург
1997

УДК 519.654(0,75.8)
Р 999

Печатается по постановлению
редакционно – издательского совета
Уральского государственного
университета им. А.М. Горького

В предлагаемом пособии приводится современное изложение метода наименьших квадратов. На основе теории псевдообращения матриц конструируется нормальное псевдорешение общей системы линейных алгебраических уравнений. Последовательно разрабатывается статистический подход, трактующий задачу отыскания неизвестного решения в русле теории оптимального оценивания (теорема Гаусса – Маркова). Значительное внимание уделяется построению оптимальных решающих процедур в рекуррентной форме, что позволяет обрабатывать большие объемы данных в реальном времени. Излагаются методы оценивания в динамических системах (фильтр Калмана – Бьюси).

Рецензенты: ОНЗАиП ИММ УрО РАН;
зав. каф. сетевых информ. систем
Урал. гос. проф.–пед. ун–та
канд. физ. – мат. наук **В.Н. Ларионов**

ISBN 5-230-06797-7

ВВЕДЕНИЕ

Вот уже двести лет метод наименьших квадратов (МНК) является одним из основных математических инструментов естествознания. Он стал широко известен после того, как К. Ф. Гауссу в 1801 г. удалось с его помощью предсказать траекторию первого открытого астероида Цереры с поразительной по тем временам точностью. По свидетельству самого Гаусса, он начал постоянно использовать этот метод, начиная с 1795 г. в работах по астрономии, а затем и в геодезических расчетах.

Исходная эвристическая идея МНК состоит в переходе от неформализованной проблемы отыскания неизвестного решения приближенной системы

$$Ax \doteq y \tag{1}$$

с прямоугольной $m \times n$-матрицей A к ясной математической задаче минимизации нормы невязки

$$\|Ax - y\| \to \min_x, \tag{2}$$

эквивалентной решению системы нормальных уравнений

$$A^{\mathsf{T}} Ax = A^{\mathsf{T}} y. \tag{3}$$

Эту идею Гаусс считал "... столь простой, что нужно скорее удивляться тому, что до нее не дошли сто лет тому назад."

Гораздо большее значение для него как математика имело обоснование метода наименьших квадратов с точки зрения теории вероятностей, а также разработка эффективных численных методов.

Исследование Гауссом связей МНК и теории вероятностей стало той основой, на которой впоследствии выросла теория оптимального оценивания. Интерпретация возмущений и ошибок наблюдений как случайных величин позволила в рамках статистического подхода ответить на многие вопросы. Так, например, теорема Гаусса – Маркова решает задачу о наилучшем выборе весовых коэффициентов в сумме квадратов невязок.

Среди последних достижений теории оптимального оценивания, получивших широкое практическое применение, следует отметить фильтр Калмана – Бьюси (1960), который рекуррентным образом формирует оценку текущего состояния динамической системы по мере поступления новых данных.

Этот фильтр можно рассматривать как обобщение предложенного Гауссом в 1821 г. рекуррентного варианта МНК, позволяющего корректировать ранее вычисленную оценку с учетом вновь поступивших дополнительных измерений без необходимости повторять все предшествующие вычисления.

Задача разработки эффективных численных методов для вырожденных случаев привела к появлению теории некорректных задач, введению операций псевдообращения и регуляризации.

Раздел пособия "МНК – детерминированная теория" посвящен изложению метода наименьших квадратов решения линейных систем в рамках детерминированного подхода. В пп. 1–3 для простейшего невырожденного случая ($\mathrm{r}(A) = n$) вводятся основные понятия и конструкции МНК, обсуждается его геометрический смысл, демонстрируется возможность рекуррентных алгоритмов.

В п. 4 рассматривается преобразование вырожденного случая к невырожденному (регуляризация). В п. 5 приводится построение МНК для общего случая. Решается задача минимизации формы второго порядка. Этот результат (теорема 5) является базовым для решения многих задач, рассматриваемых в данном пособии. Вводится понятие нормального псевдорешения, рассматривается конструкция и исследуются свойства псевдообратной матрицы. Демонстрируются (п. 6) возможности анализа системы с помощью сингулярного разложения. Приводятся (п. 7) соответствующие рекуррентные процедуры.

Раздел " Оптимальное оценивание" посвящен статистическому подходу к решению линейных систем со случайными ошибками в данных. Отличительной чертой данного пособия является стремление автора в использовании теории вероятностей ограничиться минимально необходимыми сведениями. Для понимания материала читателю достаточно знать лишь простейшие свойства математического ожидания, вводимые в п. 1 аксиоматически. Необходимые понятия ковариации и коэффициента корреляции вводятся и разъясняются по ходу решения простейшей статистической задачи.

В пп. 2,3 строится оптимальная несмещенная оценка, доказывается теорема Гаусса – Маркова. В п. 4 проводится построение оптимальной смещенной оценки. Такая оценка позволяет добиться большей точности в сравнении с несмещенной, но требует при этом дополнительные данные о матрицах вторых моментов исходного решения и помех.

В п. 5 излагается минимаксный подход, позволяющий обойтись в задании этих матриц грубыми мажорантами. Минимаксный подход указывает статистически обоснованное значение параметра регуляризации.

Раздел "Оптимальное рекуррентное оценивание" посвящен разработке рекуррентных процедур оптимального оценивания. В п. 1 доказывается основная теорема, позволяющая в п. 2 найти оптимальную оценку решения линейной системы при наличии априорной информации, в п. 3 построить рекуррентную процедуру коррекции этой оценки при получении очередного уравнения системы, в п. 4 вывести уравнения фильтра Калмана – Бьюси.

МНК–детерминированная теория

1. МНК в невырожденном случае

Рассмотрим систему линейных алгебраических уравнений

$$Ax = y, \tag{1.1}$$

где $x = (x_1, \ldots, x_n)^\top$ – неизвестный n-вектор, $y = (y_1, \ldots, y_m)^\top$ – известный m-вектор, $A = [a_{ij}]$ – заданная прямоугольная $m \times n$-матрица.

Метод наименьших квадратов решения системы (1.1) состоит в переходе от (1.1) к задаче минимизации

$$J(x) \overset{\Delta}{=} \|Ax - y\|^2 \to \min_x, \tag{1.2}$$

где $\|y\|^2 = (y, y) = \sum_{i=1}^m y_i^2$. Вектор $\hat{x}$ – решение задачи (1.2) – будем называть *МНК–решением системы* (1.1).

Для функции J(x) справедливо представление

$$J(x) = (Ax - y, Ax - y) = (Ax, Ax) - 2\,(Ax, y) + (y, y) =$$

$$= (x, A^\top Ax) - 2\,(x, A^\top y) + (y, y) = (x, Gx) - 2\,(x, b) + \alpha.$$

Следовательно, МНК сводится к минимизации формы второго порядка $J(x)$, которая складывается из квадратичной формы (x, Gx) с квадратной $n \times n$ – матрицей $G = A^\top A$, линейной формы (x, b) с n-вектором $b = A^\top y$ и константы $\alpha = (y, y)$. Поскольку $(A^\top A)^\top = A^\top A$ и

$$(x, A^\top Ax) = (Ax, Ax) \geq 0$$

при любых x, то матрица $G = A^\top A$ является симметрической и неотрицательно определенной.

Необходимое условие экстремума $\operatorname{grad} J(x) = 0$ (здесь $\operatorname{grad} J(x) = 2A^\top Ax - 2A^\top \cancel{b}\,y$) приводит к системе нормальных уравнений

$$A^\top Ax = A^\top \cancel{b}.\,y \tag{1.3}$$

1.1. Минимизация формы второго порядка с положительно определенной матрицей

Рассмотрим функцию

$$F(x) \overset{\Delta}{=} (x, Gx) - 2\,(x, b).$$

Лемма 1. *Пусть матрица G-невырожденная и $\hat{x} = G^{-1}b$. Тогда для любого n-вектора u справедливо равенство*

$$F(\hat{x} + u) = F(\hat{x}) + (u, Gu).$$

Доказательство следует из соотношений

$$F(\hat{x}+u) = (\hat{x} + u, G(\hat{x} + u)) - 2(\hat{x} + u, b) = (\hat{x}, G\hat{x}) + 2\,(u, G\hat{x}) + (u, Gu) -$$

$$-2(\hat{x}, b) - 2(u, b) = \{ \text{ т.к. } G\hat{x} = b\} = (\hat{x}, G\hat{x}) - 2(\hat{x}, b) + (u, Gu).$$

Следствие. *Пусть матрица G -положительно определенная $\Big((u, Gu) > 0$ при любом $u \neq 0$ и $(u, Gu) = 0$ лишь при $u = 0\Big)$. Тогда функция $F(x)$ достигает своего минимума на единственном векторе $\hat{x} = G^{-1}b$.*

1.2. МНК–решение и его геометрический смысл

Рассмотрим m-векторы $h_1, \ldots, h_n$ – столбцы матрицы A. Тогда $Ax = \sum_{i=1}^{n} x_i h_i$ – линейная комбинация этих векторов.

Лемма 2. *Для того чтобы матрица $A^\top A$ была положительно определенной, необходимо и достаточно, чтобы векторы $h_1, \ldots, h_n$ были линейно независимы.*

Доказательство непосредственно следует из соотношений

$$(x, A^\top Ax) = (Ax, Ax) = \|\sum_{i=1}^{n} x_i h_i\|^2.$$

Замечание. Линейная независимость столбцов $h_1, \ldots, h_n$ матрицы A означает, что $\mathrm{r}\,(A) = n$ и $m \geq n$.

Из лемм 1 и 2 вытекает следующий результат.

Теорема 1. *Пусть $\mathrm{r}\,(A) = n$. Тогда вектор $\hat{x} = \left(A^\top A\right)^{-1} A^\top y$ является единственным МНК–решением системы (1.1). Этот вектор является единственным решением системы нормальных уравнений (1.3).*

Рассмотрим геометрический смысл полученного МНК–решения. Для этого перепишем задачу (1.2) в виде

$$\left\|\sum_{i=1}^{n} x_i h_i - y\right\|^2 \to \min_{x_1, \ldots, x_n}.$$

Тогда координаты $\hat{x}_1, \ldots, \hat{x}_n$ вектора МНК-решения $\hat{x}$ системы (1.1) являются коэффициентами линейной комбинации векторов $h_1, \ldots, h_n$, наилучшим образом приближающей заданный вектор y. Хорошо известно, что в линейном подпространстве $H =< h_1, \ldots, h_n >$ ближайшим к y является вектор $\hat{y} = \text{пр}_H y$ – проекция y на H. Здесь вектор $\hat{y} - y$ должен быть ортогонален векторам $h_1, \ldots, h_n$. Требование ортогональности

$$\left(h_j, \sum_{i=1}^{n} x_i h_i - y\right) = 0, j = 1, \ldots, n$$

приводит к системе

$$\begin{aligned} (h_1, h_1)\, x_1 + (h_1, h_2)\, x_2 + \cdots + (h_1, h_n)\, x_n &= (h_1, y) \\ &\cdots \\ (h_n, h_1)\, x_1 + (h_n, h_2)\, x_2 + \cdots + (h_n, h_n)\, x_n &= (h_n, y), \end{aligned}$$

являющейся подробной записью введеной ранее системы (1.3).

Таким образом, система нормальных уравнений (1.3) имеет простой геометрический смысл, а именно задает требование ортогональности вектора $Ax - y$ векторам–столбцам матрицы A.

Определитель $\Gamma(h_1, \ldots, h_n) = \det\ [(h_i, h_j)]_{i,j=1}^{n}$ матрицы $A^\top A$ есть определитель Грама системы векторов $h_1, \ldots, h_n$.

2.Модификации МНК

2.1 МНК с весами

С произвольной симметрической положительно определенной $m \times m$-матрицей $P\,(P > 0)$ можно связать новую норму $||y||_P \overset{\Delta}{=} \sqrt{(y, Py)}$. Тогда *МНК с весовой матрицей* P представляет собой задачу минимизации

$$||Ax - y||_P^2 \to \min_x .$$

Здесь

$$||Ax - y||_P^2 = (Ax - y, P\,(Ax - y)) = \left(x, A^\top P A x\right) - 2\left(x, A^\top P y\right) + (y, Py) .$$

При этом система нормальных уравнений приобретает вид

$$A^\top P A x = A^\top P y. \tag{2.1}$$

Теорема 2. *Пусть $P > 0$ и $\operatorname{r}(A) = n$. Тогда вектор*

$$\hat{x} = \left[A^\top P A\right]^{-1} A^\top P y$$

является для системы (1.1) единственным МНК-решением с весовой матрицей P. Этот вектор является единственным решением системы нормальных уравнений (2.1).

Доказательство теоремы 2 аналогично доказательству теоремы 1. Рекомендации по выбору весовой матрицы P даются в п. 3 разд. "Оптимальное оценивание" (теорема Гаусса – Маркова) в русле статистического подхода.

2.2 МНК с линейными ограничениями

Пусть для x наряду с приближенной системой (1.1) задана точная система

$$Bx = z \tag{2.2}$$

с прямоугольной $p \times n$-матрицей B, $z = (z_1, \ldots, z_p)^\top$ – известный p-вектор. Тогда от задачи (1.2) на безусловный экстремум следует перейти к задаче на условный экстремум

$$\|Ax - y\|^2 \to \min_{x: Bx = z}. \tag{2.3}$$

Вектор $\hat{x}_1$ – решение задачи (2.3) – будем называть *МНК-решением системы (1.1) с ограничениями (2.2).*

Теорема 3. *Пусть $\operatorname{r}(A) = n, \operatorname{r}(B) = p$. Тогда матрица $B\left(A^\top A\right)^{-1} B^\top$ невырождена и система (1.1) с ограничениями (2.2) имеет единственное МНК-решение $\hat{x}_1$, связанное с $\hat{x}$ – МНК-решением системы (1.1) соотношением*

$$\hat{x}_1 = \hat{x} + \left(A^\top A\right)^{-1} B^\top \left[B\left(A^\top A\right)^{-1} B^\top\right]^{-1} (z - B\hat{x}).$$

Д о к а з а т е л ь с т в о. Очевидно, что $\hat{x}_1$–решение системы (2.2). Тогда общее решение системы (2.2) можно записать в виде

$$x = \hat{x}_1 + u = \hat{x} + d + u,$$

где u – произвольный вектор из $\ker B$, а

$$d = \left(A^\top A\right)^{-1} B^\top v, \quad v = \left[B\left(A^\top A\right) B^\top\right]^{-1} (z - B\hat{x}).$$

На основании леммы 1 для $G = A^\top A > 0$ справедливо равенство

$$J(x) = J(\hat{x}) + (d+u, G(d+u)).$$

Поскольку

$$(u, Gd) = \left(u, B^\top v\right) = (Bu, v) = \{u \in \ker B\} = 0,$$

то

$$J(x) = J(\hat{x}) + (d, Gd) + (u, Gu) = J(\hat{x}_1) + (u, Gu).$$

Таким образом, вектор $\hat{x}_1$ действительно является единственным решением задачи (2.3).

3. Рекуррентные алгоритмы МНК

Разобьем систему (1.1) на две части:

$$A_1 x \dot{=} y_{(1)}, \tag{3.1}$$

$$A_2 x \dot{=} y_{(2)}, \tag{3.2}$$

где A_1– $t \times n$-матрица, $y_{(1)}$– t-вектор, A_2– $(m-t) \times n$-матрица, $y_{(2)}$ $(m-t)$–вектор, $1 \le t \le m-1$. При этом матрица A и вектор y системы (1.1) разбиваются на блоки

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \qquad y = \begin{bmatrix} y_{(1)} \\ y_{(2)} \end{bmatrix}.$$

Пусть $\mathrm{r}(A_1) = \mathrm{r}(A) = n$. В этом случае $n \le t$. Обозначим $D = \left(A^\top A\right)^{-1}$, $D_1 = \left(A_1^\top A_1\right)^{-1}$. Рассмотрим векторы $\hat{x} = DA^\top y$ – МНК решение системы (1.1) и $\hat{x}_1 = D_1 A_1^\top y_1$ – МНК–решение системы (3.1).

Теорема 4. *Векторы $\hat{x}$ и $\hat{x}_1$ связаны соотношением*

$$\hat{x} = \hat{x}_1 + D_1 A_2^\top \left[I + A_2 D_1 A_2^\top\right]^{-1} \left(y_{(2)} - A_2 \hat{x}_1\right), \tag{3.3}$$

при этом

$$D = D_1 - D_1 A_2^\top \left[I + A_2 D_1 A_2^\top\right]^{-1} A_2 D_1. \tag{3.4}$$

Доказательство. Рассмотрим систему $A^\top Au = v$ с произвольным n-вектором v в правой части. Эта система с учетом разложения

$$A^\top A = A_1^\top A_1 + A_2^\top A_2$$

может быть записана в виде

$$\left(A_1^\top A_1 + A_2^\top A_2\right) u = v.$$

Умножив обе части этой системы слева сначала на D_1, а затем на A_2, получим системы

$$u + D_1 A_2^\top u = D_1 v, \tag{3.5}$$

$$A_2 u + A_2 D_1 A_2^\top A_2 u = A_2 D_1 v. \tag{3.6}$$

Поскольку матрица $A_2 D_1 A_2^\top$ неотрицательно определенная, то существует матрица $S = \left[I + A_2 D_1 A_2^\top\right]^{-1}$. Из (3.6) найдем вектор $A_2 u = SA_2 D_1 v$. Подставив его во второе слагаемое левой части (3.5), получим вектор $u = D_1 v - D_1 A_2^\top S A_2 D_1 v$. Поскольку $u = Dv$, то $D = D_1 - D_1 A_2^\top S A_2 D_1$. Соотношение (3.4) доказано. Соотношение (3.3) теперь проверяется непосредственными вычислениями:

$$\hat{x} = DA^\top y = \left(D_1 - D_1 A_2^\top S A_2 D_1\right)\left(A_1^\top y_{(1)} + A_2^\top y_{(2)}\right) =$$

$$= D_1 A_1^\top y_{(1)} - D_1 A_2^\top S A_2 D_1 A_1^\top y_{(1)} + D_1 A_2^\top y_{(2)} - D_1 A_2^\top S A_2 D_1 A_2^\top y_{(2)} =$$

$$= \hat{x}_1 - D_1 A_2^\top S A_2 \hat{x}_1 + D_1 A_2^\top \left[I - S\left(A_2 D_1 A_2^\top + I - I\right)\right] y_{(2)} =$$

$$= \hat{x}_1 - D_1 A_2^\top S A_2 \hat{x}_1 + D_1 A_2^\top S y_{(2)} = \hat{x}_1 + D_1 A_2^\top S\left(y_{(2)} - A_2 \hat{x}_1\right).$$

Рассмотрим отдельно случай, когда в разбиении (3.1),(3.2) система (3.2) состоит из одного последнего уравнения ($t = m-1$) системы (1.1). Тогда $A_2 = a^\top$, где $a^\top$ – последняя вектор-строка матрицы A, $y_{(2)}$ – последняя координата вектора y. При этом разбиение (3.1),(3.2) имеет вид

$$\begin{bmatrix} A_1 \\ a^\top \end{bmatrix} x = \begin{bmatrix} y_{(1)} \\ y_{(2)} \end{bmatrix},$$

а соотношения (3.3),(3.4) можно записать следующим образом:

$$\hat{x} = \hat{x}_1 + D_1 a \frac{y_{(2)} - a^\top \hat{x}_1}{1 + a^\top D_1 a}, \tag{3.7}$$

$$D = D_1 - \frac{1}{1 + a^\top D_1 a} D_1 a a^\top D_1. \tag{3.8}$$

Как видим, пересчет МНК–решения, вызванный добавлением в систему еще одного уравнения, достаточно просто осуществляется по формуле (3.7).

Используя соотношения (3.7),(3.8), легко получить следующий рекуррентный вариант МНК. Система (1.1) состоит из m уравнений

$$a_t^{\mathsf{T}} x \doteq y_t, \qquad t = 1, \ldots, m, \tag{3.9}$$

где n–вектор–строки a_t^{T} составляют матрицу A.

Рассмотрим систему

$$A_t x \doteq y^{(t)}, \tag{3.10}$$

состоящую из первых t уравнений (3.9) системы (1.1). Здесь

$$A_t = \begin{bmatrix} a_1^{\mathsf{T}} \\ \cdots \\ a_t^{\mathsf{T}} \end{bmatrix}, \quad y^{(t)} = \begin{bmatrix} y_1 \\ \cdots \\ y_t \end{bmatrix}, \quad A_m = A, \quad y^{(m)} = y\ .$$

Будем считать, что $\mathrm{r}(A) = \mathrm{r}(A_n) = n$. Это означает, что для всех t, меняющихся в диапазоне $n \le t \le m$, у матриц A_t ранг равен n и, следовательно, существуют матрицы $D_t = \left(A_t^{\mathsf{T}} A_t\right)^{-1}$ и векторы $\hat{x}_t = D_t A_t^{\mathsf{T}} y^{(t)}$ МНК–решений систем (3.10). Последовательность пар $\{\hat{x}_n, D_n\}, \ldots, \{\hat{x}_m, D_m\}$ удовлетворяет [см. (3.7),(3.8)] уравнениям

$$\hat{x}_{t+1} = \hat{x}_t + D_t a_{t+1} \frac{y_{t+1} - a_{t+1}^{\mathsf{T}} \hat{x}_t}{1 + a_{t+1}^{\mathsf{T}} D_t a_{t+1}}, \tag{3.11}$$

$$D_{t+1} = D_t - \frac{1}{1 + a_{t+1}^{\mathsf{T}} D_t a_t} D_t a_{t+1} a_{t+1}^{\mathsf{T}} D_t, \tag{3.12}$$

где $t = n, n+1, \ldots, m-1$.

Роль начальных значений в рекуррентной процедуре (3.11), (3.12) играют $\hat{x}_n = A_n^{-1} y^{(n)}$, $D_n = A_n^{-1}\left(A_n^{-1}\right)^{\mathsf{T}}$.

Пример. МНК–решение $\hat{x} = \left(\sum_{t=1}^{m} y_t\right)/m$ системы скалярных уравнений $x \doteq y_t, t = 1, \ldots, m$ может быть найдено из рекуррентных уравнений

$$\hat{x}_{t+1} = \hat{x}_t + D_t \frac{y_{t+1} - \hat{x}_t}{1 + D_t} = \left(\frac{D_t}{1 + D_t}\right) y_{t+1} + \left(\frac{1}{1 + D_t}\right) \hat{x}_t,$$

$$D_{t+1} = D_t - \frac{D_t^2}{1 + D_t} = \frac{D_t}{1 + D_t},$$

где $x_1 = y_1, D_1 = 1$.

4. Регуляризация МНК

В случае когда матрица $A^\top A$ вырождена (близка к вырожденной), от задачи (1.2) переходят к задаче

$$J_\alpha(x) \stackrel{\Delta}{=} \|Ax - y\|^2 + \alpha \|x - \bar{x}\|^2 \to \min_x \quad , \qquad \alpha > 0, \tag{4.1}$$

где $\bar{x}$–некоторый заданный n-вектор. В качестве $\bar{x}$ обычно берут какое-то известное приближение x. Если об x нет никаких сведений, то можно взять $\bar{x} = 0$.

Для функции $J_\alpha(x)$ справедливо представление

$$J_\alpha(x) = \left(x, \left(A^\top A + \alpha I\right) x\right) - 2\left(x, A^\top y + \alpha \bar{x}\right) + (y, y) + \alpha\,(\bar{x}, \bar{x}).$$

Матрица $A^\top A + \alpha I$ при $\alpha > 0$ всегда будет положительно определенной. Тогда (см. следствие к лемме 1) независимо от возможных вырождений матрицы $A^\top A$ задача (4.1) имеет единственное решение

$$\hat{x}_\alpha = \left(A^\top A + \alpha I\right)^{-1}\left(A^\top y + \alpha \bar{x}\right) = \bar{x} + \left(A^\top A + \alpha I\right)^{-1} A^\top \left(y - A\bar{x}\right). \tag{4.2}$$

При $\bar{x} = 0$

$$\hat{x}_\alpha = \left(A^\top A + \alpha I\right)^{-1} A^\top y.$$

Вектор $\hat{x}_\alpha$ называется *МНК–регуляризованным решением системы* (1.1). Скалярный коэффициент α называют *параметром регуляризации*.

Описанный здесь регуляризованный вариант МНК для системы (1.1) можно представить в форме обычного МНК для расширенной системы

$$\tilde{A}x \doteq \tilde{y}, \tag{4.3}$$

где

$$\tilde{A} = \begin{bmatrix} \beta I \\ A \end{bmatrix} - (n+m) \times n\text{–матрица}, \quad \tilde{y} = \begin{bmatrix} \beta \bar{x} \\ y \end{bmatrix} - n + m\text{–вектор},$$

$$\beta = \sqrt{\alpha}.$$

Действительно, функция

$$J_\alpha(x) = \|Ax - y\|^2 + \|\beta x - \beta \bar{x}\|^2 = \left\|\tilde{A}x - \tilde{y}\right\|^2$$

является квадратом нормы невязки для расширенной системы (4.3). Структура матрицы $\tilde{A}$ такова, что $\mathrm{r}\left(\tilde{A}\right) = n$ независимо от матрицы A. При этом вектор $\hat{x}_\alpha$ становится обычным МНК-решением системы (4.3), а вектор $\bar{x}$ можно трактовать как МНК-решение системы $\beta x \doteq \beta\bar{x}$. Тогда из теоремы 4 $\left(A_1 = \beta I, A_2 = A, D_1 = \alpha^{-1} I,\ y_{(1)} = \beta\bar{x}, y_{(2)} = y\right)$ следует равенство

$$\hat{x}_\alpha = \bar{x} + A^\top \left[\alpha I + AA^\top\right]^{-1} (y - A\bar{x}). \tag{4.4}$$

Различия (4.2) и (4.4) легко объясняются следующей леммой.

Лемма 3. *Пусть A–произвольная $m \times n$–матрица, $\alpha > 0$. Тогда*

$$A^\top \left(AA^\top + \alpha I\right)^{-1} = \left(A^\top A + \alpha I\right)^{-1} A^\top.$$

Д о к а з а т е л ь с т в о. Требуемое соотношение легко получить, если очевидное тождество $A^\top\left(AA^\top + \alpha I\right) = \left(A^\top A + \alpha I\right) A^\top$ умножить слева на $\left(A^\top A + \alpha I\right)^{-1}$, а справа – на $\left(AA^\top + \alpha I\right)^{-1}$.

Вектор $\hat{x}_\alpha$ можно вычислять рекуррентно. Рассмотрим последовательность $\hat{x}_{\alpha,t}$ – МНК–регуляризованных решений систем $A_t x \doteq y^{(t)}$, $t = 1, 2, \ldots, m$. Векторы $\hat{x}_{\alpha,t}$ являются обычными МНК–решениями расширенных систем $\tilde{A}_t x \doteq \tilde{y}^{(t)}$, где

$$\tilde{A}_t = \begin{bmatrix} \beta I \\ \ldots \\ A_t \end{bmatrix}, \quad \tilde{y}^{(t)} = \begin{bmatrix} \beta\bar{x} \\ \ldots \\ y^{(t)} \end{bmatrix}$$

Из (3.11),(3.12) следуют соотношения

$$\hat{x}_{\alpha,t+1} = \hat{x}_{\alpha,t} + D_t a_{t+1} \frac{y_{t+1} - a_{t+1}^\top \hat{x}_{\alpha,t}}{1 + a_{t+1}^\top D_t a_{t+1}}$$

$$D_{t+1} = D_t - \frac{1}{1 + a_{t+1}^\top D_t a_{t+1}} D_t a_{t+1} a_{t+1}^\top D_t,$$

здесь $D_t = \left(\tilde{A}_t^\top \tilde{A}_t\right)^{-1} = \left(A_t^\top A_t + \alpha I\right)^{-1}$, $t = 1, \ldots, m-1$; $\hat{x}_{\alpha,0} = \bar{x}$, $D_0 = \alpha^{-1} I$. Полученные рекуррентные соотношения можно выразить через матрицы: $S_t = \alpha D_t$:

$$\hat{x}_{\alpha,t+1} = \hat{x}_{\alpha,t} + S_t a_{t+1} \frac{y_{t+1} - a_{t+1}^\top \hat{x}_{\alpha,t}}{\alpha + a_{t+1}^\top S_t a_{t+1}}, \qquad \hat{x}_{\alpha,0} = \bar{x}, \tag{4.5}$$

$$S_{t+1} = S_t - \frac{1}{\alpha + a_{t+1}^\top S_t a_{t+1}} S_t a_{t+1} a_{t+1}^\top S_t, \qquad S_0 = I. \tag{4.6}$$

Статистически оптимальный вариант выбора параметра регуляризации приводится в п. 5 разд. "Оптимальное оценивание".

5. Построение МНК–решения в общем случае

Существование классического МНК–решения системы (1.1) – вектора $\bar{x} = \left(A^\top A\right)^{-1} A^\top y$ – гарантируется (см. теорему 1) требованием $\mathrm{r}(A) = n$, в условиях которого матрица $A^\top A$ становится положительно определенной, а вектор $\bar{x}$ является одновременно единственным решением задачи (1.2) и системы нормальных уравнений (1.3). При этом у $m \times n$-матрицы A количество строк должно быть не меньше количества столбцов ($m \geq n$).

Рассмотрим теперь задачу (1.2) в случае произвольной матрицы A, когда m и n – любые, а ранг $\mathrm{r} = \mathrm{r}(A)$ может лежать в пределах $0 \leq \mathrm{r} \leq \leq \min(m, n)$. Задача (1.2) сводится к минимизации формы второго порядка:

$$F(x) = (x, Gx) - 2(x, b),$$

где $G = A^\top A, b = A^\top y$. Соотношение $0 \leq J(x) = F(x) + (y, y)$ означает, что функция $F(x)$ является ограниченной снизу.

5.1 Минимизация формы второго порядка. Общий случай

Рассмотрим функцию $F(x) = (x, Gx) - 2(x, b)$ с симметрической неотрицательно определенной $n \times n$–матрицей G и n–вектором b.

Лемма 4. *Пусть $F(x)$ – ограничена снизу. Тогда для любого вектора u – решения системы $Gu = 0$ – справедливо равенство $(u, b) = 0$ $(b \perp \ker G)$.*

Доказательство. Пусть $Gu = 0$ и $(u, b) = \beta \neq 0$. Тогда $F(\mu \cdot \beta u) = = -2\mu \cdot \beta^2 \to -\infty$, при $\mu \to +\infty$, что противоречит ограниченности $F(x)$. Лемма доказана.

Решение задачи минимизации $F(x)$ опирается на следующее соотношение (спектральное разложение матрицы G):

$$G = \sum_{i=1}^{n} \lambda_i v_i v_i^\top . \tag{5.1}$$

Здесь G – произвольная симметрическая матрица, $\lambda_1, \ldots, \lambda_n$ и $v_1, \ldots, v_n$ – ее собственные значения и ортонормированная система собственных векторов: $Gv_i = \lambda_i v_i, (v_i, v_j) = \delta_{i,j}$.

Для доказательства (5.1) возьмем произвольный вектор x и разложим его по базису $v_1, \ldots, v_n$: $x = \sum_{i=1}^{n} x_i v_i, x_i = (v_i, x_i)$. Тогда $Gx = \sum_{i=1}^{n} x_i \lambda_i v_i$. Вместе с тем

$$\left[\sum_{i=1}^{n} \lambda_i v_i v_i^\top\right] x = \sum_{i=1}^{n} \lambda_i v_i v_i^\top x = \sum_{i=1}^{n} x_i \lambda_i v_i.$$

Формула (5.1) доказана.

У неотрицательно определенной матрицы G все собственные числа λ_i неотрицательны. Будем предполагать, что $\mathrm{r}\,(G) = r$ и

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > \lambda_{r+1} = \ldots = \lambda_n = 0.$$

Функция $F(x)$ для $x = \sum_{i=1}^{n} x_i v_i$ имеет вид

$$F(x) = \sum_{i=1}^{n} \lambda_i x_i^2 - 2 \sum_{i=1}^{n} \beta_i x_i,$$

где $\beta_i = (v_i, b)$. Поскольку $\lambda_{r+1} = \ldots = \lambda_n = 0$ и $\beta_{r+1} = \ldots = \beta_n = 0$ (см. лемму 4), то в действительности справедливо равенство

$$F(x) = \sum_{i=1}^{r} \lambda_i x_i^2 - 2 \sum_{i=1}^{r} \beta_i x_i = \sum_{i=1}^{r} \lambda_i \left(x_i - \frac{\beta_i}{\lambda_i}\right)^2 - \sum_{i=1}^{r} \frac{\beta_i^2}{\lambda_i}. \tag{5.2}$$

Как видим при $r < n$, вследствие вырожденности матрицы G, функция $F(x)$ не зависит от координат $x_{r+1}, \ldots, x_n$. Из (5.2) следует, что минимум $F(x)$ достигается на векторах

$$x = \sum_{i=1}^{r} \hat{x}_i v_i + \sum_{i=r+1}^{n} x_i v_i,$$

где

$$\hat{x}_i = \frac{\beta_i}{\lambda_i} \qquad (i = 1, \ldots, r),$$

а $x_{r+1}, \dots, x_n$ – произвольные числа.

Координаты $\hat{x}_1, \dots, \hat{x}_r$ задают вектор

$$\hat{x} = \sum_{i=1}^{r} \hat{x}_i v_i = Qb \; , \; Q = \sum_{i=1}^{r} \lambda_i^{-1} v_i v_i^{\top} . \qquad (5.3)$$

При этом линейные комбинации $u = \sum_{i=r+1}^{n} x_i v_i$ составляют подпространство $\mathbf{U}$ всех решений системы $Gu = 0 \, (U = \ker G)$. Отметим, что $\dim \mathbf{U} = n - \mathbf{r}$. Все векторы, на которых достигается минимум $F(x)$, составляют аффинное подпространство $\mathbf{X} = \hat{x} + \mathbf{U}$. Вектор $\hat{x}$ в подпространстве $\mathbf{X}$ занимает особое положение. Действительно, благодаря ортогональности $\hat{x}$ всем векторам $u \in \mathbf{U}$ справедливо соотношение

$$\|\hat{x} + u\|^2 = (\hat{x} + u, \hat{x} + u) = (\hat{x}, \hat{x}) + 2\,(\hat{x}, u) + (u, u) = \|\hat{x}\|^2 + \|u\|^2,$$

из которого следует, что $\hat{x}$ – единственный вектор, имеющий в $\mathbf{X}$ наименьшую норму. В заключение найдем, что

$$\min_x F(x) = -\sum_{i=1}^{r} \lambda_i^{-1} \beta_i^2 = -\sum_{i=1}^{r} \lambda_i^{-1} b^{\top} v_i v_i^{\top} b = -b^{\top} Q b.$$

Итогом этих рассуждений является

Теорема 5. *Минимум ограниченной снизу функции*

$$F(x) = (x, Gx) - 2\,(x, b)$$

достигается на векторах

$$x = \hat{x} + u, \qquad (5.4)$$

где $\hat{x} = Qb$, $Q = \sum_{i=1}^{r} \lambda_i^{-1} v_i v_i^{\top}$, *$u$ – произвольный вектор из подпространства* $\mathbf{U} = \ker G$. *В подпространстве* $\mathbf{X} = \hat{x} + \mathbf{U}$ *единственным вектором, имеющим наименьшую норму, является вектор* $\hat{x}$. *При этом* $\min_x F(x) = F(\hat{x}) = -b^{\top} Q b$.

5.2 Нормальное псевдорешение

Рассмотрим систему (1.1) с произвольной прямоугольной $m \times n$–матрицей A. Пусть r – ее ранг : $0 \le \mathrm{r} \le m, n$. Вектор x, минимизирующий $J(x) \triangleq \|Ax - y\|^2$, будем называть *псевдорешением* системы (1.1).

Рассмотрим величину $J_0 = \min_x J(x)$ и множество $\mathbf{X} = \{x | J(x) = J_0\}$ — всех псевдорешений системы (1.1). Наименьший по норме вектор из $\mathbf{X}$ будем называть *нормальным псевдорешением* системы (1.1).

Напомним, что $J(x) = (x, Gx) - 2(x, b) + (y, y)$, где $G = A^\top A$, $b = A^\top y$. Благодаря теореме 5 справедлив следующий результат:

Теорема 6. *Система (1.1) при любой матрице A и правой части y, имеет единственное нормальное псевдорешение*

$$\hat{x} = Ly, \tag{5.5}$$

где

$$L = \left(\sum_{i=1}^{r} \lambda_i^{-1} v_i v_i^\top \right) A^\top, \tag{5.6}$$

$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$ — положительные собственные числа матрицы $A^\top A$; $v_1, \ldots, v_r$ — соответствующая им ортонормированная система собственных векторов. Множество всех псевдорешений $\mathbf{X}$ есть аффинное подпространство векторов вида $x = \hat{x} + u$, где $u \in \ker\left(A^\top A\right)$.

Лемма 5. *Подпространство всех решений системы $A^\top A u = 0$ совпадает с подпространством всех решений системы $Au = 0$.*

Доказательство непосредственно следует из соотношений

$$\|Au\|^2 = (Au, Au) = \left(u, A^\top A u\right).$$

Поскольку

$$A^\top A \hat{x} = \sum_{i=1}^{r} \lambda_i^{-1} \left(A^\top A\right) v_i v_i^\top A^\top y = \sum_{i=1}^{r} v_i v_i^\top A^\top y =$$

$$= \left\{ v_i^\top A^\top \left(A v_i\right)^\top = 0 \text{ при } i = r+1, \ldots, n \text{ (см. лемму 5)} \right\} =$$

$$= \left(\sum_{i=1}^{n} v_i v_i^\top \right) A^\top y = A^\top y,$$

то система нормальных уравнений (1.3) всегда совместна. Все ее решения составляют подпространство $\mathbf{X}$. Таким образом, нормальное псевдорешение системы (1.1) есть наименьшее по норме решение системы нормальных уравнений (1.3).

5.3 Псевдообратная матрица и ее свойства

Пусть A – произвольная прямоугольная $m \times n$–матрица. Матрицу L размерности $n \times m$, задающую нормальное псевдорешение (5.5), (5.6) системы (1.1), будем называть *псевдообратной* к A и обозначать A^+.

Рассмотрим псевдообращение симметрической матрицы. Пусть S – произвольная симметрическая $n \times n$–матрица ранга r,

$$\lambda_1 \neq 0, \ldots, \lambda_r \neq 0, \lambda_{r+1} = \ldots = \lambda_n = 0$$

– собственные числа, $s_1, \ldots, s_r, s_{r+1}, \ldots, s_n$ – ортонормированная система соответствующих им собственных векторов. Тогда S имеет следующее спектральное разложение:

$$S = \sum_{i=1}^{r} \lambda_i s_i s_i^\top . \tag{5.7}$$

Вследствие ортонормированности векторов s_i справедливо равенство

$$S^\top S = \sum_{i=1}^{r} \lambda_i^2 s_i s_i^\top ,$$

представляющее собой спектральное разложение матрицы $S^\top S = S^2$. Используя формулу (5.6), находим

$$S^+ = \left[\sum_{i=1}^{r} \lambda_i^{-2} s_i s_i^\top \right] S^\top = \sum_{i=1}^{r} \lambda_i^{-1} s_i s_i^\top . \tag{5.8}$$

Таким образом, псевдообращение симметрической матрицы в терминах спектральных разложений [см.(5.7),(5.8)] производится по следующему правилу: собственные векторы не меняются, а ненулевые собственные числа заменяются на обратные величины. В частности, при $r = n$ (S–невырожденная)

$$S^+ = \sum_{i=1}^{n} \lambda_i^{-1} s_i s_i^\top = S^{-1}$$

Теперь для вектора $\hat{x}$, минимизирующего функцию $F(x) = (x, Gx) - -2(x, b)$ (см. теорему 5), справедлива краткая запись $\hat{x} = G^+ b$. Из (5.6) и (5.8) следует формула

$$A^+ = \left[A^\top A\right]^+ A^\top , \tag{5.9}$$

сводящая псевдообращение прямоугольной матрицы A общего вида к псевдообращению симметрической матрицы $A^\top A$.

Свойства псевдообратной матрицы

1. $A^+ = A^\top \left[AA^\top\right]^+$
2. $\left(A^\top\right)^+ = \left(A^+\right)^\top$
3. $\left(A^+\right)^+ = A$
4. $AA^+A = A$
5. $A^+AA^+ = A^+$
6. $\left(AA^+\right)^\top = AA^+$
7. $\left(AA^+\right)^2 = AA^+$
8. $\left(A^+A\right)^\top = A^+A$
9. $\left(A^+A\right)^2 = A^+A$
10. $\lim_{\alpha\searrow 0}\left[A^\top A + \alpha I\right]^{-1} A^\top =$
    $= \lim_{\alpha\searrow 0} A^\top \left(AA^\top + \alpha I\right)^{-1} = A^+$
11. $\lim_{\alpha\searrow 0}\alpha\left[A^\top A + \alpha I\right]^{-1} = I - A^+A$
12. $A^+Ax = \text{пр}_H x$, $H = < a_1, \ldots, a_m >$,
    где $a_1, \ldots, a_m$ –столбцы матрицы $A^\top$.

Д о к а з а т е л ь с т в о. Пусть

$$A^\top A = \sum_{i=1}^{r} \lambda_i v_i v_i^\top \; - \tag{5.10}$$

спектральное разложение матрицы $A^\top A$. Тогда, благодаря ортонормированности векторов $v_1, \ldots, v_n$, справедливы равенства

$$A^+A = \left[A^\top A\right]^+ A^\top A = \sum_{i=1}^{r} v_i v_i^\top. \tag{5.11}$$

Отметим, что для единичной матрицы I справедливо разложение $I = \sum_{i=1}^{n} v_i v_i^\top$. Поэтому $A^+A \neq I$ при $r \neq n$. Как видим, матрица A^+A является симметрической и $\left(A^+A\right)^2 = \sum_{i=1}^{r} v_i v_i^\top = A^+A$. Свойства 8,9 доказаны. Умножим равенство (5.11) слева на A и, учитывая лемму 5, получим

$$AA^+A = \sum_{i=1}^{r} Av_i v_i^\top = \sum_{i=1}^{n} Av_i v_i^\top = A. \tag{5.12}$$

Свойство 4 доказано. Транспонируя равенство (5.12) с учетом свойства 8, получим

$$A^+AA^\top = A^\top. \tag{5.13}$$

Подставляя в (5.13) вместо матрицы A матрицу $A^\top$, получим равенство

$$\left(A^\top\right)^+ A^\top A = A, \tag{5.14}$$

из которого находим

$$A^\top = \left(A^\top A\right)^\top \left[\left(A^\top\right)^+\right]^\top = A^\top A \left[\left(AA^\top\right)^+ A\right]^\top = A^\top AA^\top \left(AA^\top\right)^+ .$$

Подставляя найденное $A^\top$ в выражение (5.9) для A^+ с учетом (5.13), получим

$$A^+ = \left(A^\top A\right)^+ A^\top = \left(A^\top A\right)^+ A^\top AA^\top \left(AA^\top\right)^+ = A^\top \left(AA^\top\right)^+ .$$

Свойство 1 доказано. Умножим равенство (5.13) справа на $\left(AA^\top\right)^+$, получим

$$A^+AA^\top \left(AA^\top\right)^+ = A^\top \left(AA^\top\right)^+ ,$$

откуда с учетом свойства 1 следует равенство

$$A^+AA^+ = A^+ .$$

Свойство 5 доказано. Подставляя в 1 вместо A матрицу $A^\top$, получим

$$\left(A^\top\right)^+ = A\left(A^\top A\right)^+ .$$

С другой стороны, $\left(A^+\right)^\top = \left[\left(A^\top A\right)^+ A^\top\right]^\top = A\left(A^\top A\right)^+$. Таким образом, $\left(A^\top\right)^+ = \left(A^+\right)^\top$. Свойство 2 доказано. Подставляя в равенство $B^+ = \left(B^\top B\right)^+ B^\top$ матрицу $B = A^+ = A^\top \left(AA^\top\right)^+$, получим

$$\left(A^+\right)^+ = \left[\left(AA^\top\right)^+ \left(AA^\top\right)\left(AA^\top\right)^+\right]^+ \left(AA^\top\right)^+ A. \tag{5.15}$$

Поскольку $\left(AA^\top\right)^+ \left(AA^\top\right)\left(AA^\top\right)^+ = \left(AA^\top\right)^+$ (см. свойство 5), а $\left(\left(AA^\top\right)^+\right)^+ = AA^\top$ [для симметрических матриц это следует из (5.11)], то из (5.15) следует $\left(A^+\right)^+ = AA^\top \left(AA^\top\right)^+ A = \{$ свойство 1 $\} = AA^+A = \{$ свойство 4 $\} = A$. Свойство 3 доказано. Из равенств

$$AA^+ = A\left(A^\top A\right)^+ A^\top$$

и

$$\left(AA^\top\right)^2 = A\left(A^\top A\right)^+ A^\top A\left(A^\top A\right)^+ A^\top = A\left(A^\top A\right)^+ A^\top$$

следуют свойства 6 и 7.

Из леммы 3 и разложения (5.10) следуют равенства

$$R_\alpha \stackrel{\Delta}{=} \left(A^\top A + \alpha I\right)^{-1} A^\top = A^\top \left(AA^\top + \alpha I\right)^{-1} = \sum_{i=1}^{n} (\lambda_i + \alpha)^{-1} v_i v_i^\top A^\top .$$

Поскольку $Av_{r+1} = \ldots = Av_n = 0$ (лемма 5), то

$$R_\alpha = \sum_{i=1}^{r} (\lambda_i + \alpha)^{-1} v_i v_i^\top A^\top .$$

Переходя к пределу по $\alpha \searrow 0$, получим равенства из свойства 10. Из свойства 10 следует, что нормальное псевдорешение $\bar{x}$ является пределом МНК–регуляризованного ($\bar{x} = 0$) решения $\bar{x}_\alpha$ при $\alpha \searrow 0$.

Из (5.10) следуют равенства

$$\alpha \left(A^\top A + \alpha I\right)^{-1} = \sum_{i=1}^{n} \alpha (\lambda_i + \alpha)^{-1} v_i v_i^\top =$$

$$= \sum_{i=1}^{r} \alpha (\lambda_i + \alpha)^{-1} v_i v_i^\top + I - A^+ A.$$

Переходя к пределу по $\alpha \searrow 0$, получим равенство из свойства 11.

Докажем свойство 12. Пусть

$$A^\top = [a_1, \ldots, a_m] .$$

Справедливы равенства

$$A^+ A a_j = \left\{(5.10) \text{ и } a_j = A^\top e_j\right\} = \sum_{i=1}^{r} v_i v_i^\top A^\top e_j =$$

$$= \{ \text{ лемма } 5\} = \sum_{i=1}^{n} v_i v_i^\top A^\top e_j = A^\top e_j = a_j .$$

Как видим, матрица $A^+ A$ оставляет все векторы $a_1, \ldots, a_m$ на месте. Все пространство n–мерных векторов X разлагается на прямую сумму ортогональных подпространств H и $\ker\left(A^+ A\right)$:

$$X = H + \ker\left(A^+ A\right) .$$

Следовательно, матрица A^+A есть проектор на подпространство II. Все свойства доказаны.

Псевдообратная матрица единственным образом определяется уравнениями Пенроуза (см. свойства 4, 5, 6, 8)

$$AXA \stackrel{a}{=} A, \quad XAX \stackrel{b}{=} X, \quad (XA)^\top \stackrel{c}{=} XA, \quad (AX)^\top \stackrel{d}{=} AX.$$

Действительно, пусть матрицы X и Y удовлетворяют этим уравнениям. Тогда справедлива следующая цепочка равенств:

$$X \stackrel{b}{=} XAX \stackrel{d}{=} X(AX)^\top = XX^\top A^\top \stackrel{a}{=} XX^\top A^\top Y^\top A^\top =$$

$$= X(AX)^\top (AY)^\top \stackrel{d}{=} XAXAY \stackrel{b}{=} XAY \stackrel{b}{=} XAYAY \stackrel{c}{=}$$

$$\stackrel{c}{=} (XA)^\top (YA)^\top Y = A^\top X^\top A^\top Y^\top Y \stackrel{a}{=}$$

$$\stackrel{a}{=} A^\top Y^\top Y = (YA)^\top Y \stackrel{c}{=} YAY \stackrel{a}{=} Y.$$

Отсюда, в частности, следует единственность псевдообратной матрицы, несмотря на извесный произвол в выборе векторов $v_1, \ldots, v_n$.

Примеры.

1. Пусть α – число, тогда

$$\alpha^+ = \begin{cases} 0, & \alpha = 0; \\ \frac{1}{\alpha}, & \alpha \neq 0. \end{cases}$$

2. Пусть $a = (a_1, \ldots, a_m)^\top$ — вектор-столбец, $b = (b_1, \ldots, b_n)$ — вектор-строка. Тогда

$$a^+ = \left(a^\top a\right)^+ a^\top = \begin{cases} (0, \ldots, 0), & a = 0; \\ \frac{a^\top}{a^\top a}, & a \neq 0; \end{cases}$$

$$b^+ = b^\top \left(bb^\top\right)^+ = \begin{cases} (0, \ldots, 0)^\top, & b = 0; \\ \frac{b^\top}{bb^\top}, & b \neq 0. \end{cases}$$

3. Пусть A– $m \times n$–матрица. Если $\mathrm{r}(A) = n$, то $A^+ = \left(A^\top A\right)^{-1} A^\top$. Если $\mathrm{r}(A) = m$, то $A^+ = A^\top \left(AA^\top\right)^{-1}$.

4. Пусть $m \times n$–матрица A ранга r имеет вид

$$A = \begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix},$$

где R – квадратная невырожденная $r \times r$- матрица. Тогда

$$A^+ = \begin{bmatrix} R^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

6. Сингулярное разложение

Рассмотрим матрицу $A^\top A$, ее собственные числа

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0 = \lambda_{r+1} = \ldots = \lambda_n,$$

где $\mathrm{r} = \mathrm{r}\left(A^\top A\right) = \mathrm{r}\left(A\right)$ и соответствующий им ортонормированный набор собственных векторов $v_1, v_2, \ldots, v_r, v_{r+1}, \ldots, v_n$.

Лемма 6. *Пусть $\lambda \neq 0$ и v – собственное число и нормированный собственный вектор матрицы $A^\top A$. Тогда вектор $u = \frac{1}{\sqrt{\lambda}}Av$ является нормированным собственным вектором матрицы $A^\top A$ для того же собственного значения λ.*

Доказательство. Справедливость леммы следует из равенств

$$\|u\|^2 = (u, u) = \frac{1}{\lambda}(Av, Av) = \frac{1}{\lambda}\left(v, A^\top Av\right) = (v, v) = \|v\|^2 = 1,$$

$$AA^\top u = \frac{1}{\sqrt{\lambda}}AA^\top Av = \sqrt{\lambda}Av = \frac{\lambda}{\sqrt{\lambda}}Av = \lambda u.$$

Следствие. *Матрицы $A^\top A$ и $AA^\top$ имеют одинаковый набор ненулевых собственных чисел.*

Величины

$$\rho_1 = \sqrt{\lambda_1}, \ \ldots, \rho_r = \sqrt{\lambda_r}\,, \ \rho_{r+1} = 0, \ldots, \rho_k = 0 \ \left(\mathrm{r} = \mathrm{r}\left(A\right), k = \min(n, m)\right)$$

называются *сингулярными числами* матрицы A.

Ортонормированные системы собственных векторов $v_1, \ldots, v_n$ и $u_1, \ldots, u_m$ матриц $A^\top A$ и $AA^\top$ называются *сингулярными базисами* матрицы A.

Рассмотрим $m \times m$- матрицу $U = [u_1, \ldots, u_m]$, $n \times n$- матрицу $V = [v_1, \ldots, v_n]$ и $m \times n$ матрицу Λ, у которой на главной диагонали стоят сингулярные числа, а вне диагонали все элементы равны нулю.

Теорема 7. *Для произвольной $m \times n$ матрицы A справедливы представления*

$$A = \sum_{i=1}^{r} \rho_i u_i v_i^\top, \qquad A^+ = \sum_{i=1}^{r} \rho_i^{-1} v_i u_i^\top; \tag{6.1}$$

$$A = U\Lambda V^\top, \qquad A^+ = V\Lambda^+ U^\top. \tag{6.2}$$

Доказательство. Цепочки равенств

$$A = \left\{ I = \sum_{i=1}^{n} v_i v_i^\top \right\} = A \sum_{i=1}^{n} v_i v_i^\top = \{Av_{r+1} = \ldots = Av_n = 0\} =$$

$$= \sum_{i=1}^{r} Av_i v_i^\top = \left\{u_i = \rho_i^{-1} Av_i\right\} = \sum_{i=1}^{r} \rho_i u_i v_i^\top,$$

$$A^+ = \left(A^\top A\right)^+ A^\top = \sum_{i=1}^{r} \rho_i^{-2} v_i v_i^\top A^\top = \sum_{i=1}^{r} \rho_i^{-1} v_i u_i^\top$$

доказывают разложения (6.1). Разложения (6.2) есть матричная запись разложений (6.1).

Рассмотрим систему

$$y = Ax + \xi. \tag{6.3}$$

Используя спектральное разложение (6.2), перепишем (6.3) в виде

$$y = U\Lambda V^\top x + \xi.$$

После умножения слева на $U^\top$ получим

$$U^\top y = \Lambda V^\top x + U^\top \xi.$$

Сделав замену

$$U^\top y = q, \quad V^\top x = p, \quad U^\top \xi = \eta,$$

получим систему

$$q = \Lambda p + \eta,$$

которая для координат $q_i = (u_i, y)$, $p_i = (v_i, x)$, $\eta_i = (u_i, \xi)$ имеет вид

$$\begin{cases} q_1 &= \rho_1 p_1 + \eta_1, \\ & \ldots \\ q_r &= \rho_r p_r + \eta_r, \\ q_{r+1} &= \eta_{r+1}, \\ & \ldots \\ q_m &= \eta_m. \end{cases}$$

Таким образом, система (6.3) при помощи разложения по сингулярным базисам распадается на независимые уравнения.

Полученное сингулярное разложение исходной системы позволяет сделать следующие выводы. Вектор данных q (а следовательно, и y) никак не связан с координатами $p_{r+1}, \ldots, p_n$. Это означает, что данная система не содержит никакой информации относительно этих координат, а поэтому о значениях этих координат ничего сказать нельзя. Для оставшихся координат $p_1, \ldots, p_r$ вектора неизвестных p из первых r независимых уравнений можно указать приближения

$$\hat{p}_i = \frac{q_i}{\rho_i}, \qquad i = 1, \ldots, r. \tag{6.5}$$

При этом координаты $q_{r+1}, \ldots, q_m$ вектора данных связаны лишь с помехами и, следовательно, являются бесполезными для оценки p. Вектор $\hat{x} = \sum_{i=1}^{r} \hat{p}_i v_i$ является [см. (5.5), (5.6)] нормальным псевдорешением системы $Ax \doteq y$.

Приближения $\hat{p}_i$ имеют следующие ошибки:

$$\hat{p}_i - p_i = \frac{\eta_i}{\rho_i}.$$

Как видим, координаты $\hat{p}_i$, отвечающие малым сингулярным числам ρ_i, являются неустойчивыми к ошибкам η_i. В этих обстоятельствах применяется *дискретная регуляризация*.

Для сингулярных чисел задается некоторая нижняя граница — число ε. По ε находится значение дискретного параметра s такого, что

$$\rho_1 \geq \ldots \geq \rho_s \geq \varepsilon > \rho_{s+1} \geq \ldots \geq \rho_r > 0.$$

Отсекая малые сингулярные числа, получим s–регуляризованное решение

$$\hat{x}^{(s)} = \sum_{i=1}^{s} \hat{p}_i v_i = A_s^+ y,$$

где $A_s = U\Lambda_s V^\top$, а матрица Λ_s получена из Λ заменой малых сингулярных чисел $\rho_{s+1}, \ldots, \rho_r$ — нулями. При этом $A_s^+ = V\Lambda_s^+ U^\top$.

7. Рекуррентное вычисление нормального псевдорешения

Рассмотрим наряду с системой (1.1) последовательность систем (3.10). Соответствующие нормальные псевдорешения $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_m$ могут быть найдены рекуррентно.

Теорема 8. *Пусть A — $m \times n$ матрица и $\mathrm{r}(A) = m \le n$. Тогда для векторов $\hat{x}_t = A_t^+ y^{(t)}$ — нормальных псевдорешений систем $A_t x = y^{(t)}$ и матриц $P_t \triangleq I - A_t^+ A_t$, $t = 1, \ldots, m$, справедливы рекуррентные соотношения*

$$\hat{x}_{t+1} = \hat{x}_t + \frac{P_t a_{t+1}}{a_{t+1}^\top P_t a_{t+1}} \left(y_{t+1} - a_{t+1}^\top \hat{x}_t \right), \qquad \hat{x}_0 = 0; \qquad (7.1)$$

$$P_{t+1} = P_t - \frac{P_t a_{t+1} a_{t+1}^\top P_t}{a_{t+1}^\top P_t a_{t+1}}, \qquad P_0 = I \qquad (7.2)$$

Д о к а з а т е л ь с т в о. Поскольку $\mathrm{r}(A) = m$, то все строки $a_t^\top$, $t = 1, \ldots, m$, матрицы A линейно независимы и, следовательно, $a_{t+1} \notin \langle a_1, \ldots, a_t \rangle$. Поскольку P_t — матрица оператора проектирования на подпространство $\langle a_1, \ldots, a_t \rangle^\perp$, то

$$a_{t+1}^\top P_t a_{t+1} > 0.$$

Положим в (4.5),(4.6) $x = 0$, тогда

$$\bar{x}_{\alpha,t} = \left(A_t^\top A_t + \alpha I \right)^{-1} A_t y^{(t)},$$

$$S_{\alpha,t} = \alpha \left(A_t^\top A_t + \alpha I \right)^{-1}.$$

Переходя в (4.5), (4.6) к пределу при $\alpha \to 0$ с учетом свойств 10 и 11, получим соотношения (7.1),(7.2).

Оптимальное оценивание

1. Введение

1.1. Необходимые сведения из теории вероятностей и математической статистики

Из всего обширного материала, традиционно излагаемого в курсе теории вероятностей и математической статистики, нам для дальнейших построений понадобятся следующие сведения.

Пусть x — случайная величина. Ее наиболее простой детерминированной характеристикой является ***математическое ожидание*** $\mathbf{E}(x)$.

Свойства математического ожидания

1. Если случайная величина $x \geq 0$, то $\mathbf{E}(x) \geq 0$.
2. Если x и y — случайные, а a и b — детерминированные величины, то
$$\mathbf{E}(ax + by) = a\,\mathbf{E}(x) + b\,\mathbf{E}(y).$$
3. Если a — детерминированная величина, то $\mathbf{E}(a) = a$.

Другой важной характеристикой случайной величины x является ее ***дисперсия***:
$$\mathbf{D}(x) \stackrel{\triangle}{=} \mathbf{E}\,(x - \mathbf{E}\,x)^2 .$$
Если a — детерминированная величина, то $\mathbf{D}(a) = 0$.

Если $\mathbf{E}(x - y)^2 = 0$, то будем говорить, что случайные величины x и y совпадают ***в среднем квадратичном*** ($x \stackrel{\text{с.к}}{=} y$) и в дальнейшем не будем их различать.

Для пары случайных величин x и y определяется ***ковариация***
$$\operatorname{cov}(x, y) \stackrel{\triangle}{=} \mathbf{E}\left\{(x - \mathbf{E}\,x)(y - \mathbf{E}\,y)\right\}.$$
Отметим, что $\operatorname{cov}(x, x) = \mathbf{D}(x)$, $\operatorname{cov}(x, y) = \mathbf{E}(xy) - \mathbf{E}(x)\,\mathbf{E}(y)$. Функция $\operatorname{cov}(x, y)$ естественным образом возникает при решении следующей задачи.

Пусть x и y — случайные величины. Рассмотрим задачу наилучшего приближения величины x линейной комбинацией $ay + b$, где a и b

– детерминированные величины. В качестве меры погрешности такого приближения возьмем функцию $J(a,b) \stackrel{\Delta}{=} \mathbf{E}\left(x - (ay+b)\right)^2$. Задача наилучшего среднеквадратичного приближения x линейной комбинацией $ay+b$ сводится к задаче минимизации

$$J(a,b) \to \min_{a,b}.$$

Для функции $J(a,b)$, используя свойства 2 и 3 математического ожидания, легко получить следующее представление:

$$J(a,b) = \mathbf{E}\left((x-ay)-b\right)^2 = b^2 + 2\left(a\,\mathbf{E}(y) - \mathbf{E}(x)\right)b + \mathbf{E}\left((x-ay)^2\right).$$

Поскольку

$$\frac{\partial J}{\partial b} = 2b + 2\left(a\,\mathbf{E}(y) - \mathbf{E}(x)\right),$$

то для каждого a минимум по b функции $J(a,b)$ достигается при

$$b(a) = \mathbf{E}(x) - a\,\mathbf{E}(y)$$

При этом

$$\begin{aligned} J(a,b(a)) &= \mathbf{E}\left((x-ay) - \mathbf{E}(x) + a\,\mathbf{E}(y)\right)^2 \\ &= \mathbf{E}\left((x - \mathbf{E}(x)) - a\,(y - \mathbf{E}(y))\right)^2 \\ &= \mathrm{cov}(x,x) - 2a\mathrm{cov}(x,y) + a^2\mathrm{cov}(y,y). \end{aligned}$$

Из уравнения

$$\frac{dJ(a,b(a))}{da} = -2\mathrm{cov}(x,y) + 2a\mathrm{cov}(y,y) = 0,$$

при $\mathrm{cov}(y,y) \neq 0$ найдем оптимальное значение $a_0 = \mathrm{cov}(x,y)/\mathrm{cov}(y,y)$ и

$$\begin{aligned} \min_{a,b} J(a,b) = J(a_0, b(a_0)) &= \mathrm{cov}(x,x) - \frac{\mathrm{cov}^2(x,y)}{\mathrm{cov}(y,y)} \\ &= \mathrm{cov}(x,x)\left[1 - \frac{\mathrm{cov}^2(x,y)}{\mathrm{cov}(x,x)\mathrm{cov}(y,y)}\right]. \end{aligned}$$

Таким образом, для случайной величины x при $\mathrm{cov}(y,y) \neq 0$ наилучшим среднеквадратичным приближением в классе линейных комбинаций $ay+b$ будет

$$\hat{x} = \mathbf{E}(x) + \frac{\mathrm{cov}(x,y)}{\mathrm{cov}(y,y)}\left(y - \mathbf{E}(y)\right),$$

при этом
$$\mathbf{E}\,(x-\hat{x})^2 = \operatorname{cov}(x,x) - \frac{\operatorname{cov}^2(x,y)}{\operatorname{cov}(y,y)}. \tag{1.1}$$
Величину
$$\mathbf{K}(x,y) = \frac{\operatorname{cov}(x,y)}{\sqrt{\operatorname{cov}(x,x)\operatorname{cov}(y,y)}}$$
называют *коэффициентом корреляции x и y*.

Погрешность (1.1) выражается через $\mathbf{K}$ следующим образом:
$$\mathbf{E}\,(x-\hat{x})^2 = \operatorname{cov}(x,x)\left(1-\mathbf{K}^2\right).$$
Значения $|\mathbf{K}|$ лежат в интервале $0 \le |\mathbf{K}| \le 1$. Отметим крайние случаи:

(a) $|\mathbf{K}| = 1$.
Здесь $\mathbf{E}\,(x-\hat{x})^2 = 0$, т.е. $\hat{x} \overset{\text{с.к.}}{=} x$.

(b) $\mathbf{K} = 0$.
Здесь $\operatorname{cov}(x,y) = 0$ и наилучшим приближением для x будет $\hat{x} = \mathbf{E}(x)$, т.е. сведения об y ничего не дают для уточнения x.

В случае $\operatorname{cov}(x,y) = 0$ величины x и y называют *некоррелированными*. Для некоррелированных величин x и y справедливы равенства
$$\mathbf{E}(xy) = \mathbf{E}(x)\,\mathbf{E}(y), \quad \mathbf{D}(x+y) = \mathbf{D}(x) + \mathbf{D}(y).$$

1.2. Случайные векторы и их детерминированные характеристики

Пусть $x = (x_1,\ldots,x_n)^\top$ и $y = (y_1,\ldots,y_m)^\top$ — векторы, координаты которых есть случайные числа. Будем называть *математическим ожиданием вектора x* детерминированный вектор
$$\mathbf{E}(x) \overset{\Delta}{=} (\mathbf{E}(x_1),\ldots,\mathbf{E}(x_n))^\top,$$
ковариационной матрицей векторов x и y — детерминированную матрицу
$$\operatorname{cov}(x,y) = [\operatorname{cov}(x_i,y_j)]_{i,j=1}^{n,m},$$
матрицей смешанных вторых моментов векторов x и y — матрицу
$$\mathbf{E}\left(xy^\top\right) = [\mathbf{E}(x_iy_j)]_{i,j=1}^{n,m}.$$

Если $\operatorname{cov}(x, y) = 0$, то векторы x и y будем называть *некоррелированными*. При этом

$$\mathbf{E}\left(xy^{\top}\right) = \mathbf{E}(x)\,\mathbf{E}(y)^{\top}. \tag{1.2}$$

Пусть $a = (a_1, \ldots, a_n)^{\top}$ и $b = (b_1, \ldots, b_m)^{\top}$ — детерминированные векторы. Из свойств 2, 3 математического ожидания для скалярных величин следуют равенства для векторов

$$\mathbf{E}\left(a^{\top} x\right) = a^{\top}\,\mathbf{E}(x),\ \mathbf{E}\left(a^{\top} x y^{\top} b\right) = a^{\top}\,\mathbf{E}\left(xy^{\top}\right) b. \tag{1.3}$$

1.3 Оценки и их свойства

Пусть дана серия наблюдений $y_1, y_2, \ldots, y_m$, связанная каким-то образом с искомой величиной x. Для неизвестного x по данным $y_1, \ldots, y_m$ требуется построить приближение x. Естественно взять

$$\hat{x} = \varphi(y_1, \ldots, y_m) = \varphi(y). \tag{1.4}$$

Функцию $\varphi(y)$ будем называть *оценкой*.

Свойства оценок

1. Линейность.
   Оценка (1.4) называется *линейной*, если функция $\varphi(y)$ — линейна.

2. Несмещенность.
   Оценка (1.4) называется *несмещенной*, если $\mathbf{E}(\hat{x}) = \mathbf{E}(x)$.

3. Оптимальность.
   Рассмотрим класс оценок (1.4), где $\varphi \in \Phi$. Каждой функции $\varphi \in \Phi$ соответствует своя оценка $\hat{x}_{\varphi} = \varphi(y)$. В качестве меры погрешности оценки $\hat{x}_{\varphi}$ величины x возьмем $\mathbf{E}\left(x - \hat{x}_{\varphi}\right)^2$. Оценка $\hat{x}_{\varphi_0}$ называется *оптимальной в классе* Φ, если

$$\mathbf{E}\left(x - \hat{x}_{\varphi_0}\right)^2 = \inf_{\varphi \in \Phi} \mathbf{E}\left(x - \hat{x}_{\varphi}\right)^2.$$

Если x — детерминированная величина, то несмещенность оценки $\hat{x}_{\varphi}$ означает, что $\mathbf{E}\,\hat{x}_{\varphi} = x$ и $\mathbf{E}\left(x - \hat{x}_{\varphi}\right)^2 = \mathbf{E}\left(\hat{x}_{\varphi} - \mathbf{E}\,\hat{x}_{\varphi}\right)^2 = \mathbf{D}\left(\hat{x}_{\varphi}\right)$. В этом случае задача построения оптимальной оценки совпадает с задачей построения оценки, имеющей минимальную дисперсию.

Примеры. Пусть наблюдения $y_1, \ldots, y_m$ связаны с x уравнениями $y_t = x + \xi_t$, $t = 1, \ldots, m$. Рассмотрим оценки

$$\hat{x}_1 = \frac{1}{2}\left[\max_t y_t + \min_t y_t\right], \quad \hat{x}_2 = \frac{1}{m}\sum_{t=1}^{m} y_t, \quad \hat{x}_3 = \sum_{t=1}^{m} h_t y_t,$$

h_t — детерминированные параметры. Здесь оценка $\hat{x}_1$ является нелинейной, а оценки $\hat{x}_2$ и $\hat{x}_3$ — линейными. Значения $\hat{x}_2$ и $\hat{x}_3$ связаны с x соотношениями

$$\hat{x}_2 = x + \frac{1}{m}\sum_{t=1}^{m}\xi_t, \quad \hat{x}_3 = \left(\sum_{t=1}^{m} h_t\right) x + \sum_{t=1}^{m} h_t \xi_t.$$

Пусть $\mathbf{E}(\xi_t) = 0$, $t = 1, \ldots, m$. Тогда оценка $\hat{x}_2$ является несмещенной. Оценка $\hat{x}_3$ будет несмещенной, если параметры h_t удовлетворяют условию

$$\sum_{t=1}^{m} h_t = 1. \tag{1.5}$$

При этом

$$\hat{x}_3 = x + \sum_{t=1}^{m} h_t \xi_t, \quad \mathbf{E}\,(x - \hat{x}_3)^2 = \mathbf{E}\left(\sum_{t=1}^{m} h_t \xi_t\right)^2 =$$

$$= \sum_{i,j=1}^{m} \mathbf{E}\,(\xi_i \xi_j)\, h_i h_j = h^\top V h,$$

где

$$h = (h_1, \ldots, h_m)^\top,$$

$V = \mathbf{E}(\xi\xi^\top)$ — матрица смешанных вторых моментов вектора $\xi = (\xi_1, \ldots, \xi_m)^\top$.

Соотношение (1.5) определяет класс несмещенных оценок вида $\hat{x}_3$, а квадратичная форма $J(h) \triangleq h^\top V h$ задает погрешности таких оценок. В итоге задача выбора параметров h оптимальной несмещенной оценки $\hat{x}_3$ сводится к следующей детерминированной задаче:

$$h^\top V h \to \min_{e^\top h = 1}, \tag{1.6}$$

где $e^\top = (1, \ldots, 1)$ — m-мерный вектор. Задача (1.6) для $m = 2$ и $\det V \neq 0$ выглядит следующим образом:

$$J(h_1, h_2) = \mathbf{E}\left(\xi_1^2\right) h_1^2 + 2\,\mathbf{E}\,(\xi_1\xi_2)\, h_1 h_2 + \mathbf{E}\left(\xi_2^2\right) h_2^2 \to \min_{h_1 + h_2 = 1}$$

и имеет простой геометрический смысл.

На прямой, задаваемой уравнением $h_1 + h_2 = 1$, требуется найти точку пересечения (касания) $h = (h_1, h_2)$ с наименьшим эллипсом из семейства эллипсов линий уровня для квадратичной формы $J(h_1, h_2)$.

В случае, когда помехи ξ_1 и ξ_2 некоррелированы ($\mathbf{E}(\xi_1\xi_2) = 0$) , оси эллипсов совпадают с осями координат. Пусть $\mathbf{E}\left(\xi_1^2\right) = \sigma_1$, $\mathbf{E}\left(\xi_2^2\right) = \sigma_2$, тогда

$$J(h_1, h_2) = \sigma_1 h_1^2 + \sigma_2 h_2^2,$$

$$\min_{h_1+h_2=1} \left\{\sigma_1 h_1^2 + \sigma_2 h_2^2\right\} = \min_{h_1} \left\{\sigma_1 h_1^2 + \sigma_2 (1 - h_1)^2\right\} = \frac{\sigma_1\sigma_2}{\sigma_1 + \sigma_2},$$

$$h_1 = \frac{\sigma_2}{\sigma_1 + \sigma_2}\ , \ h_2 = \frac{\sigma_1}{\sigma_1 + \sigma_2}.$$

2. Оптимальное оценивание решения линейной системы со случайными ошибками в данных

2.1. Линейная оценка и ее погрешность

Рассмотрим систему

$$y = Ax + \xi, \tag{2.1}$$

где x неизвестный n-вектор; y m-вектор данных, содержащий m вектор случайных ошибок ξ; A детерминированная $m \times n$-матрица Предполагается, что векторы x и ξ являются некоррелированными и $\mathbf{E}(\xi) = 0$. Нас интересует скалярная величина z, связанная с x соотношением

$$z = c^{\mathsf{T}} x, \tag{2.2}$$

где c детерминированный n-вектор. В качестве оценки неизвестного z возьмем

$$\hat{z} = h^{\mathsf{T}} y, \tag{2.3}$$

где $h^{\mathsf{T}} = (h_1, \ldots, h_m)$ m-мерный вектор детерминированных параметров. Для ошибки оценки (2.3) справедливо равенство

$$z - \hat{z}(h) = c^{\mathsf{T}} x - h^{\mathsf{T}} (Ax + \xi) = \left(c - A^{\mathsf{T}} h\right)^{\mathsf{T}} x - h^{\mathsf{T}} \xi.$$

Далее, учитывая детерминированность c, A и h и свойства математического ожидания 2, 3, получим

$$\mathbf{E}\left(z - \hat{z}(h)\right) = \left(c - A^{\top} h\right)^{\top} \mathbf{E}(x) - h^{\top}\,\mathbf{E}(\xi)$$
$$\mathbf{E}\left(z - z(h)\right)^2 = \left(c - A^{\top} h\right)^{\top} \mathbf{E}\left(xx^{\top}\right)\left(c - A^{\top} h\right) - $$
$$- 2\left(c - A^{\top} h\right)^{\top} \mathbf{E}\left(x\xi^{\top}\right) h + h^{\top}\,\mathbf{E}\left(\xi\xi^{\top}\right) h\,.$$

Поскольку x и ξ — некоррелированы, а $\mathbf{E}(\xi) = 0$, то [см.(1.2)] $\mathbf{E}\left(x\xi^{\top}\right) = \mathbf{E}(x)\,\mathbf{E}(\xi)^{\top} = 0$. В результате получаем следующие равенства:

$$\mathbf{E}\left(z - \hat{z}(h)\right) = \left(c - A^{\top} h\right)^{\top} \mathbf{E}(x) \tag{2.4}$$

$$\mathbf{E}\left(z - z(h)\right)^2 = \left(c - A^{\top} h\right)^{\top} X \left(c - A^{\top} h\right) + h^{\top} V h, \tag{2.5}$$

где

$$X = \mathbf{E}\left(xx^{\top}\right),\ V = \mathbf{E}\left(\xi\xi^{\top}\right). \tag{2.6}$$

2.2. Построение оптимальной несмещенной оценки

Будем говорить, что величина $z = c^{\top} x$ *допускает несмещенную оценку*, если найдется вектор h такой, что $\mathbf{E}\left(z(h)\right) = \mathbf{E}(z)$.

Требование несмещенности оценки (2.3) приводит [см.(2.4)] к условию

$$\left(c - A^{\top} h\right)^{\top} \mathbf{E}(x) = 0.$$

Здесь предполагается, что $\mathbf{E}(x)$ — неизвестный вектор, который, вообще говоря, может принимать любые значения. (Это предположение, в частности, продиктовано желанием охватить случай, когда неизвестный вектор x является детерминированным: $\mathbf{E}(x) = x$.) В этих обстоятельствах требование несмещенности приводит к условию

$$A^{\top} h = c. \tag{2.7}$$

Таким образом, в множестве линейных оценок (2.3) подмножество несмещенных оценок величины z составляют оценки с параметрами h, удовлетворяющими системе (2.7). Параметры всех несмещенных оценок образуют множество $\mathcal{H} = \left\{h \mid A^{\top} h = c\right\}$. Величина z допускает несмещенную оценку тогда и только тогда, когда $\mathcal{H} \neq \oslash$.

Пусть $A^{\top} = [a_1, \ldots, a_m]$, где $a_1, \ldots, a_m$ — n-векторы-столбцы матрицы $A^{\top}$. Тогда систему (2.7) можно переписать иначе:

$$\sum_{i=1}^{m} h_i a_i = c.$$

Для того чтобы величина $z = c^\top x$ допускала несмещенную оценку, необходимо и достаточно, чтобы $c \in < a_1, \ldots, a_m >$

Величина $z = c^\top x$ допускает несмещенную оценку при любом c, тогда и только тогда, когда $\mathsf{r}(A) = n$.

Говорят, что вектор x *допускает несмещенную оценку*, если каждая его координата $x_i = c_i^\top x$, где $c_i^\top = (0, \ldots, 1, \ldots, 0)$, допускает несмещенную оценку.

Вектор x допускает несмещенную оценку тогда и только тогда, когда $\mathsf{r}(A) = n$.

Пусть $\mathsf{r}(A) = n$. Тогда у системы (2.7) с правыми частями $c = e_i$, $i = 1, 2, \ldots, n$, существуют решения $h^{(1)}, \ldots, h^{(n)}$ $(A^\top h^{(i)} = c_i)$. При этом $\hat{x}_i = h^{(i)\top} y$ будут несмещенными оценками координат x_i. Рассмотрим $m \times n$-матрицу $H = \left[h^{(1)} \ldots h^{(n)}\right]$. Эта матрица удовлетворяет уравнению $A^\top H = E$ и задает n вектор оценки $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)^\top = H^\top y$.

Замечание. Требование несмещенности является весьма сильным. В некоторых случаях оно однозначно определяет оценку. Действительно, пусть $m = n = \mathsf{r}(A)$ (матрица A квадратная и невырожденная). Тогда система $A^\top h = E$ имеет единственное решение $H = \left(A^{-1}\right)^\top$ и единственно возможным вектором несмещенной оценки будет $\hat{x} = A^{-1} y$ — классическое решение.

В классе несмещенных оценок погрешность (2.5) не зависит от X и равна

$$\mathbf{E}\,(z - \hat{z}(h))^2 = h^\top V h. \tag{2.8}$$

В результате построение оптимальной несмещенной оценки сводится к задаче на условный экстремум

$$h^\top V h \longrightarrow \min_{h: A^\top h = c} \tag{2.9}$$

Теорема 9. *Пусть* $\mathsf{r}(A) = n$, $\mathsf{r}(V) = m$. *Тогда для неизвестной величины z из (2.2) по данным y из (2.1) единственной оптимальной несмещенной оценкой в классе оценок (2.3) будет*

$$\hat{z} = c^\top \left[A^\top V^{-1} A\right]^{-1} A^\top V^{-1} y. \tag{2.10}$$

Ее погрешность равна

$$\mathbf{E}\,(z - \hat{z})^2 = c^\top \left[A^\top V^{-1} A\right]^{-1} c. \tag{2.11}$$

При этом оптимальной несмещенной оценкой всего вектора x будет

$$\hat{x} = \left[A^\top V^{-1} A\right]^{-1} A^\top V^{-1} y. \tag{2.12}$$

Погрешность $\hat{x}$ равна

$$\mathbf{E}(x-\hat{x})(x-\hat{x})^\top = \left[A^\top V^{-1} A\right]^{-1}. \qquad (2.13)$$

Д о к а з а т е л ь с т в о. Множество $\mathcal{H}$ состоит из векторов вида

$$h = h_0 + \bar{h},$$

где h_0 – произвольное решение однородной системы $A^\top h_0 = 0$, а $\bar{h}$ – частное решение неоднородной системы (2.7). Возьмем

$$\bar{h} = V^{-1} A \left[A^\top V^{-1} A\right]^{-1} c$$

(существование обратных матриц следует из условий $\mathrm{r}(A) = n$, $\mathrm{r}(V) = m$). Поскольку

$$\bar{h}^\top V h_0 = c^\top \left[A^\top V^{-1} A\right]^{-1} A^\top V^{-1} V h_0 = 0,$$

то

$$h^\top V h = \left(h_0 + \bar{h}\right)^\top V \left(h_0 + \bar{h}\right) = h_0^\top V h_0 + 2\bar{h}^\top V h_0 + \bar{h}^\top V \bar{h} =$$

$$= h_0^\top V h_0 + \bar{h}^\top V \bar{h},$$

откуда, благодаря положительной определенности матрицы V, следует равенство

$$\min_{h_0} h^\top V h = \bar{h}^\top V \bar{h}^\top = c^\top \left[A^\top V^{-1} A\right]^{-1} c,$$

при этом минимум достигается на единственном векторе $h_0 = 0$. Таким образом, единственной оптимальной несмещенной оценкой будет $\hat{z} = \bar{h}^\top y$. Соотношения (2.12),(2.13) легко получить из (2.10),(2.11), если взять $\hat{x} = (\hat{z}_1, \ldots, \hat{z}_n)^\top$, где $\hat{z}_i$ – оптимальные несмещенные оценки величин $z_i = c_i^\top x = x_i$. Теорема доказана.

Следствие. В случае $V = vI$ оптимальной несмещенной оценкой является вектор

$$\hat{x} = \left[A^\top A\right]^{-1} A^\top y$$

с погрешностью

$$\mathbf{E}(x-\hat{x})(x-\hat{x})^\top = v\left[A^\top A\right]^{-1}.$$

3. Теорема Гаусса – Маркова

Теорема Гаусса – Маркова устанавливает связь между статистическим подходом разд. "Оптимальное оценивание", трактующим задачу решения линейной системы как некоторую задачу оптимального оценивания, и детерминированным подходом разд. "МНК – детерминированная теория", приводящим к МНК-решениям.

Теорема 10 (Гаусса – Маркова). *Пусть* $\mathrm{r}\,(A) = n$. *В случае, когда помехи* $\xi_1, \ldots, \xi_m$, *составляющие вектор* ξ, *являются некоррелированными и имеют* $\mathbf{E}(\xi_i^2) = v > 0$, *оптимальной несмещенной оценкой* x *по данным* $y = Ax + \xi$ *является МНК-решение* $\hat{x} = \left[A^\top A\right]^{-1} A^\top y$ *системы* $Ax \doteq y$.

В общем случае, когда помехи $\xi_1, \ldots, \xi_m$ *коррелированы, оптимальной несмещенной оценкой будет МНК-решение*

$$\hat{x} = \left[A^\top PA\right]^{-1} A^\top Py$$

с весовой матрицей $P = V^{-1}$.

Эта утверждение, непосредственно вытекающее из теоремы 1, представляет собой обоснование МНК с точки зрения теории вероятностей.

4. Построение оптимальной смещенной оценки

Если отказаться от дополнительного требования несмещенности, то построение оптимальной оценки сводится [см. (2.5),(2.6)] к задаче на безусловный экстремум:

$$J(h) \triangleq \left(c - A^\top h\right)^\top X \left(c - A^\top h\right) + h^\top V h \longrightarrow \min_h . \tag{4.1}$$

Для функции $J(h)$ справедливо представление

$$J(h) = (h, Gh) - 2(h, b) + \alpha, \tag{4.2}$$

где $G = AXA^\top + V$, $b = AXc$, $\alpha = c^\top Xc$.

Из теоремы 5 разд. "МНК – детерминированная теория" следует, что минимум функции $J(h)$ достигается на векторах

$$h = G^+ b + u, \qquad u \in \ker G,$$

и равен $J \triangleq \min J(h) = \alpha - b^\top G^+ b$. Соответствующие оценки имеют вид

$$z_u = \hat{z} + u^\top y, \qquad \hat{z} = b^\top G^+ y = c^\top X A^\top \left[A X A^\top + V\right]^+ y,$$

при этом

$$J = \mathbf{E}\,(z - \hat{z})^2 = c^\top \left[X - X A^\top \left[A X A^\top + V\right]^+ A X\right] c.$$

Поскольку $\mathbf{E}\left(y y^\top\right) = \mathbf{E}\left\{(Ax + \xi)\left(x^\top A^\top + \xi^\top\right)\right\} = A X A^\top + V = G$ и $Gu = 0$, то $\mathbf{E}\left(u^\top y\right)^2 = \mathbf{E}\left\{u^\top y y^\top u\right\} = u^\top\, \mathbf{E}\left\{y y^\top\right\} u = u^\top G y = 0$. Поскольку $u^\top y \overset{\text{с.к.}}{=} 0$, то при любом $u \in \ker G$ справедливо равенство $z_u \overset{\text{с.к.}}{=} \hat{z}$, означающее, что оценка $\hat{z}$ является единственной. В итоге справедлива

Теорема 11. *Единственной оптимальной оценкой величины $z = c^\top x$ в классе $z(h) = h^\top y$ по данным y из (2.1) является*

$$\hat{z} = c^\top X A^\top \left[A X A^\top + V\right]^+ y. \tag{4.3}$$

Ее погрешность равна

$$\mathbf{E}\,(z - \hat{z})^2 = c^\top \left[X - X A^\top \left[A X A^\top + V\right]^+ A X\right] c. \tag{4.4}$$

При этом оптимальной оценкой всего вектора x будет

$$\hat{x} = X A^\top \left[A X A^\top + V\right]^+ y. \tag{4.5}$$

Погрешность $\hat{x}$ равна

$$\mathbf{E}\,(x - \hat{x})(x - \hat{x})^\top = X - X A^\top \left[A X A^\top + V\right] A X. \tag{4.6}$$

Замечание. Отказ от требования несмещенности расширяет возможности оптимизации, хотя и требует при этом дополнительных сведений матрицу $X = \mathbf{E}\left(x x^\top\right)$. Пусть $\hat{z}_{\text{о.н.о}}$ — оптимальная несмещенная оценка, а $\hat{z}_{\text{о.с.о.}}$ — оптимальная смещенная оценка. По построению ясно, что $\mathbf{E}\,(z - \hat{z}_{\text{о.с.о}})^2 \le \mathbf{E}\,(z - \hat{z}_{\text{о.н.о.}})^2$. Погрешность $\hat{z}_{\text{о.с.о.}}$ всегда ограничена, как бы ни вырождалась матрица A. Действительно, справедливо неравенство $\mathbf{E}\,(z - \hat{z}_{\text{о.с.о.}})^2 \le c^\top X c$. Погрешность $\hat{z}_{\text{о.н.о.}}$ может принимать сколь угодно большие значения. Пусть, например, $V = vI$, c

нормированный собственный вектор матрицы $A^\top A$, соответствующий минимальному собственному значению λ. Тогда

$$\mathbf{E}\left(z-\hat{z}_{\text{о.н.о.}}\right)^2 = vc^\top \left[A^\top A\right]^{-1} c = \frac{v}{\lambda}$$

и при $\lambda \to 0$ погрешность $\mathbf{E}\left(z-\hat{z}_{\text{о.н.о}}\right)^2 \to \infty$.

Величина смещения оценки $z_{\text{о.с.о}}$ равна

$$\mathbf{E}\left(z-\hat{z}_{\text{о.с.о.}}\right) = c^\top \left[I - XA^\top\left[AXA^\top + V\right]^+ A\right]\mathbf{E}(x).$$

В одномерном случае, когда

$$y = ax + \xi,\ \mathbf{E}\left(\xi^2\right) = v,\ \mathbf{E}\left(x^2\right) = r,\ \hat{x}_{\text{о.с.о}} = \frac{ar}{a^2r+v}y,$$

смещение равно

$$\mathbf{E}\left(x-\hat{x}_{\text{о.с.о.}}\right) = \frac{v}{a^2r+v}\,\mathbf{E}(x).$$

5. Минимаксная оценка

Матрица $X = \mathbf{E}\left(xx^\top\right)$, используемая в оптимальной оценке (4.5), редко бывает известна точно. Например, в одномерном случае, когда x — детерминированная положительная величина, требование задать $X = \mathbf{E}\left(x^2\right)$ — это то же самое, что задать сам x. Здесь оптимальная оценка теряет всякий практический смысл. Вместе с тем для искомого вектора x часто бывает известно неравенство $(x,x) = \|x\|^2 \le k$.

Лемма 7. *Скалярное неравенство $(x,x) \le k$ эквивалентно матричному неравенству $xx^\top \le kI$.*

Д о к а з а т е л ь с т в о. Пусть $(x,x) = x^\top x \le k$. Тогда при любом векторе a справедливы соотношения

$$a^\top xx^\top a = (a,x)^2 \le (x,x)(a,a) \le k(a,a) = a^\top (kI)\, a,$$

из которых следует неравенство $xx^\top \le kI$.

Пусть $xx^\top \le kI$. Тогда справедливы соотношения

$$(x,x)^2 = x^\top xx^\top x \le kx^\top x = k(x,x),$$

из которых следует неравенство $(x,x) \le k$. Лемма доказана.

Рассмотрим более общую ситуацию. Пусть для неизвестных матриц $X = \mathbf{E}\left(xx^{\top}\right)$, $V = \mathbf{E}\left(\xi\xi^{\top}\right)$ заданы мажоранты

$$X \leq \bar{X},\ V \leq \bar{V}. \tag{5.1}$$

В этих обстоятельствах в качестве погрешности оценки $\bar{z}(h) = h^{\top}y$ следует взять

$$J(h) = \max_{X \leq \bar{X},\ V \leq \bar{V}} \mathbf{E}\left(z - \bar{z}(h)\right)^2 \tag{5.2}$$

и строить *минимаксную оценку (ММ оценку)*, решая задачу

$$J(h) \longrightarrow \min_h .$$

Из (2.5) следует, что

$$J(h) = \left(c - A^{\top}h\right)^{\top} \bar{X} \left(c - A^{\top}h\right) + h^{\top}\bar{V}h.$$

Тогда (см. теорему 11) ММ оценкой ($\hat{x}_{\text{м.м.о.}}$) неизвестного вектора x будет

$$\hat{x}_{\text{м.м.о.}} = \bar{X}A^{\top}\left[A\bar{X}A^{\top} + \bar{V}\right]^{+} y.$$

При этом для погрешности оценки $\hat{x}_{\text{м.м.о.}}$ справедливо неравенство

$$\mathbf{E}\left(x - \hat{x}_{\text{м.м.о.}}\right)\left(x - \hat{x}_{\text{м.м.о.}}\right)^{\top} \leq \bar{X} - \bar{X}A^{\top}\left[A\bar{X}A^{\top} + \bar{V}\right]^{+} A\bar{X}.$$

Рассмотрим случай, когда для норм неизвестных векторов x и ξ системы (2.1) заданы мажоранты $\|x\|^2 \leq k$ и $\|\xi\|^2 \leq \sigma$. Тогда, взяв $\bar{X} = kI$, $\bar{V} = \sigma I$, получим оценку

$$\hat{x}_{\text{м.м.о.}} = A^{\top}\left[AA^{\top} + \frac{\sigma}{k}I\right]^{-1} y =$$

$$= \{\text{см.лемму 3 разд."МНК -- детерминированная теория"}\} =$$

$$= \hat{x}_{\text{м.м.о.}} = \left[A^{\top}A + \frac{\sigma}{k}I\right]^{-1} Ay.$$

Как видим, найденная ММ оценка совпадает с МНК регуляризованным решением [см. (4.2) того же раздела] системы $Ax \doteq y$. При этом $\bar{x} = 0$, $\alpha = \frac{\sigma}{k}$. Следует подчеркнуть, что минимаксный подход позволяет получить здесь статистически обоснованный вариант выбора параметра регуляризации.

Для погрешности ММ оценки справедливы соотношения

$$\mathbf{E}\left(x - \hat{x}_{\text{м.м.о.}}\right)\left(x - \hat{x}_{\text{м.м.о.}}\right)^{\top} \leq k\left[I - A^{\top}\left(AA^{\top} + \frac{\sigma}{k}I\right)^{-1} A\right] =$$

$$= k\left[I - \left(A^{\mathsf T} A + \frac{\sigma}{k}\right)^{-1} A^{\mathsf T} A\right] = \sigma\left[A^{\mathsf T} A + \frac{\sigma}{k} I\right]^{-1}.$$

При $k \to \infty$ (априорные сведения об x неограниченно ухудшаются) ММ-оценка сходится (см. свойство 10 псевдообратных матриц) к вектору $\bar{x} = A^{+}y$ нормального псевдорешения.

Оптимальное рекуррентное оценивание

1. Основная теорема

Рассмотрим вектор $\vartheta = \begin{bmatrix} x \\ y \end{bmatrix}$, составленный из неизвестного n-мерного вектора x и известного m мерного вектора y. Будем предполагать, что для ϑ задана априорная оценка $\bar{\vartheta} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$ и корреляционная матрица $D = \mathbf{E}\left\{\left(\vartheta - \bar{\vartheta}\right)\left(\vartheta - \bar{\vartheta}\right)^{\top}\right\}$ ошибки $\vartheta - \bar{\vartheta}$. Матрица D имеет следующую блочную структуру:

$$D = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix} \qquad \begin{aligned} D_{xx} &= \mathbf{E}\left\{(x - \bar{x})(x - \bar{x})^{\top}\right\} \\ D_{xy} = D_{yx}^{\top} &= \mathbf{E}\left\{(x - \bar{x})(y - \bar{y})^{\top}\right\} \end{aligned}$$

$$D_{yy} = \mathbf{E}\left\{(y - \bar{y})(y - \bar{y})^{\top}\right\}.$$

Нас интересует скалярная величина

$$z = c^{\top} x, \tag{1.1}$$

где c – заданный детерминированный n-вектор. Для неизвестного z рассмотрим класс оценок

$$\bar{z}(h) = c^{\top}\bar{x} + h^{\top}(y - \bar{y}). \tag{1.2}$$

Теорема 12. *Единственной оптимальной оценкой величины (1.1) в классе (1.2) является*

$$\hat{z} = c^{\top}\bar{x} + c^{\top}D_{xy}D_{yy}^{+}(y - \bar{y}). \tag{1.3}$$

Ее погрешность равна

$$\mathbf{E}(z - \hat{z})^2 = c^{\top}\left[D_{xx} - D_{xy}D_{yy}^{+}D_{yx}\right]c. \tag{1.4}$$

При этом оптимальной оценкой всего вектора x будет

$$\hat{x} = \bar{x} + D_{xy}D_{yy}^{+}(y - \bar{y}).$$

Погрешность $\hat{x}$ равна

$$\mathbf{E}(x - \hat{x})(x - \hat{x})^{\top} = D_{xx} - D_{xy}D_{yy}^{+}D_{yx}.$$

Д о к а з а т е л ь с т в о. Для ошибки $z - \hat{z}(h)$ справедливы равенства

$$z - \hat{z}(h) = c^\top x - c^\top x - h^\top (y - y) = c^\top (x - x) - h^\top (y - y) =$$

$$= \begin{bmatrix} c \\ -h \end{bmatrix}^\top (\vartheta - \vartheta),$$

$$(z - \hat{z}(h))^2 = \begin{bmatrix} c \\ -h \end{bmatrix}^\top (\vartheta - \vartheta)(\vartheta - \vartheta)^\top \begin{bmatrix} c \\ -h \end{bmatrix},$$

$$\mathbf{E}(z - \hat{z}(h))^2 = \begin{bmatrix} c \\ -h \end{bmatrix}^\top \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix} \begin{bmatrix} c \\ -h \end{bmatrix} =$$

$$= c^\top D_{xx} c - c^\top D_{xy} h - h^\top D_{yx} c + h^\top D_{yy} h =$$

$$= h^\top D_{yy} h - 2h^\top D_{yx} c + c^\top D_{xx} c \triangleq J(h).$$

Решением задачи $J(h) \to \min_h$ является (теорема 5 разд. "МНК детерминированная теория") множество векторов

$$h = D_{yy}^+ D_{yx} c + u, \qquad u \in \ker D_{yy},$$

и значение

$$J \triangleq \min_h J(h) = c^\top D_{xx} c - c^\top D_{xy} D_{yy}^+ D_{yx} c.$$

Как видим, оценка (1.3) действительно является оптимальной и для ее погрешности справедливо равенство (1.4). Единственность оптимальной оценки (1.3) следует из равенств

$$\mathbf{E}\left(u^\top (y - y)\right)^2 = \mathbf{E}\left(u^\top (y - y)(y - y)^\top u\right) = u^\top D_{yy} u = 0.$$

Теорема доказана.

Следствие. Пусть для матрицы D известна лишь мажоранта D, $D \leq D$. Тогда минимаксной оценкой вектора x будет

$$\hat{x}_{\text{м.м.о}} = x + D_{xy} D_{yy}^+ (y - y)$$

При этом

$$\mathbf{E}(x - \hat{x}_{\text{м.м.о.}})(x - \hat{x}_{\text{м.м.о.}})^\top \leq D_{xx} - D_{xy} D_{yy}^+ D_{yx}.$$

С помощью доказанной теоремы легко выводятся уравнения (4.3), (4.4) разд. "Оптимальное оценивание". Эта теорема, позволяя охватить более общий случай (см. п. 2 разд. "Оптимальное рекуррентное оценивание"), послужит основой при выводе рекуррентных формул пп. 3,4 данного раздела.

2. Оптимальная оценка решения линейной системы при наличии априорной информации

Пусть для неизвестного x системы

$$y = Ax + \xi \tag{2.1}$$

имеется априорная оценка $\bar{x}$. Предполагается, что векторы $x - \bar{x}$ и ξ некоррелированы и заданы матрицы

$$\Delta \triangleq \mathbf{E}\,(x - \bar{x})(x - \bar{x})^{\top} = D_{xx},$$

$$V = E\left(\xi\xi^{\top}\right).$$

Взяв $\bar{y} = A\bar{x}$, получим

$$D_{xy} = \Delta A^{\top},\; D_{yy} = A^{\top}\Delta A + V.$$

Тогда по теореме 12 оптимальной оценкой неизвестного x будет вектор

$$\hat{x} = \bar{x} + \Delta A^{\top}\left[A\Delta A^{\top} + V\right]^{+}(y - A\bar{x}) \tag{2.2}$$

с погрешностью

$$\mathbf{E}\,(x - \hat{x})(x - \hat{x})^{\top} = \Delta - \Delta A^{\top}\left[A\Delta A^{\top} + V\right]^{+} A\Delta. \tag{2.3}$$

3. Рекуррентная процедура решения линейной системы

Систему (2.1) распишем по уравнениям:

$$y_t = a_t^{\top} x + \xi_t, \qquad t = 1, \ldots, m,$$

где

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \quad A = \begin{bmatrix} a_1^{\top} \\ \vdots \\ a_m^{\top} \end{bmatrix}, \quad \xi = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_m \end{bmatrix}.$$

Будем предполагать, что координаты $\xi_1, \ldots, \xi_m$ попарно некоррелированы и $\mathbf{E}\left(\xi_i^2\right) = v_i$, т.е. $V = \mathbf{E}\left(\xi\xi^{\mathsf{T}}\right) = \operatorname{diag}\left[v_1, \ldots, v_m\right]$.

Рассмотрим подсистему $y^{(t)} = A_t x + \xi^{(t)}$ системы (2.1), где

$$y^{(t)} = \begin{bmatrix} y_1 \\ \vdots \\ y_t \end{bmatrix}, \quad A_t = \begin{bmatrix} a_1^{\mathsf{T}} \\ \vdots \\ a_t^{\mathsf{T}} \end{bmatrix}, \quad \xi^{(t)} = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_t \end{bmatrix};$$

$$V_t = \mathbf{E}\left(\xi^{(t)}\xi^{(t)\mathsf{T}}\right) = \operatorname{diag}\left[v_1, \ldots, v_t\right].$$

Решением задачи оптимального оценивания x по данным $y^{(t)}$ при наличии априорной оценки $\bar{x}$ будет (см. теорему 12) вектор

$$\hat{x}_t = \bar{x} + \Delta A_t^{\mathsf{T}}\left[A_t \Delta A_t^{\mathsf{T}} + V_t\right]^{+}\left(y^{(t)} - A_t \bar{x}\right)$$

с погрешностью $\Delta_t = \mathbf{E}\left(x - \hat{x}_t\right)\left(x - \hat{x}_t\right)^{\mathsf{T}}$. Для оценок $\hat{x}_t$ выведем рекуррентную формулу. Для этого, следуя общей схеме раздела 1, введем вектор $\vartheta_{t+1} = \begin{bmatrix} x \\ y_{t+1} \end{bmatrix}$. Здесь $\hat{x}_t$, играя роль априорной оценки, позволяет найти $\bar{\vartheta}_{t+1} = \begin{bmatrix} \hat{x}_t \\ a_{t+1}^{\mathsf{T}}\hat{x}_t \end{bmatrix}$. Тогда

$$D_{t+1} \triangleq \mathbf{E}\left(\vartheta_{t+1} - \bar{\vartheta}_{t+1}\right)\left(\vartheta_{t+1} - \bar{\vartheta}_{t+1}\right)^{\mathsf{T}} =$$

$$= \begin{bmatrix} \Delta_t & \Delta_t a_{t+1} \\ a_{t+1}^{\mathsf{T}}\Delta_t & a_{t+1}^{\mathsf{T}}\Delta_t a_{t+1} + v_{t+1} \end{bmatrix}$$

Из теоремы 12 следует, что оптимальной оценкой будет вектор

$$\hat{x}_{t+1} = \hat{x}_t + \Delta_t a_{t+1}\left(a_{t+1}^{\mathsf{T}}\Delta_t a_{t+1} + v_{t+1}\right)^{+}\left(y_{t+1} - a_{t+1}^{\mathsf{T}}\hat{x}_t\right) \qquad (3.1)$$

с погрешностью

$$\Delta_{t+1} = \Delta_t - \Delta_t a_{t+1}\left(a_{t+1}^{\mathsf{T}}\Delta_t a_{t+1} + v_{t+1}\right)^{+} a_{t+1}^{\mathsf{T}}\Delta_t. \qquad (3.2)$$

Начальными данными для полученной процедуры оптимального оценивания (3.1),(3.2) являются $\hat{x}_0 = \bar{x}$, $\Delta_0 = \Delta$

4. Оптимальное оценивание в динамических системах. Фильтр Калмана – Бьюси

Рассмотрим динамическую систему

$$x_{t+1} = A_t x_t + \xi_t, \tag{4.1}$$

$$y_t = C_t x_t + \eta_t, \tag{4.2}$$

где x_t — n-мерный вектор неизвестных состояний; y_t — m-мерный вектор наблюдений; A_t и C_t — известные матрицы размерности $n \times n$ и $m \times n$; ξ_t и η_t — некоррелированные векторы случайных возмущений размерности n и m с характеристиками

$$\mathbf{E}(\xi_t) = 0\,,\ \mathbf{E}\left(\xi_t \xi_t^{\mathsf{T}}\right) = V_t\,,\ \mathbf{E}(\eta_t) = 0\,,\ \mathbf{E}\left(\eta_t \eta_t^{\mathsf{T}}\right) = W_t\,.$$

Для начального состояния x_0 системы (4.1) известна априорная оценка — вектор $\bar{x}_0$, не коррелированный с ξ_t и η_t, с матрицей погрешности $\Delta_0 = \mathbf{E}\,(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^{\mathsf{T}}$.

Пусть для x_t к моменту t $(t = 1, 2, \ldots)$ по заданным $y_0, y_1, \ldots, y_{t-1}$ известна оптимальная оценка $\hat{x}_t$ и ее погрешность

$$\Delta_t = \mathbf{E}\,(x_t - \hat{x}_t)(x_t - \hat{x}_t)^{\mathsf{T}}\,.$$

Нас интересует оптимальная оценка следующего состояния x_{t+1}, учитывающая дополнительное данное y_t. В соответствии с общей схемой составим вектор $\vartheta = \begin{bmatrix} x_{t+1} \\ y_t \end{bmatrix}$. В качестве априорной оценки для ϑ возьмем $\bar{\vartheta} = \begin{bmatrix} A_t \hat{x}_t \\ C_t \hat{x}_t \end{bmatrix}$. Здесь

$$D = \mathbf{E}\left(\vartheta - \bar{\vartheta}\right)\left(\vartheta - \bar{\vartheta}\right)^{\mathsf{T}} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix},$$

где $D_{xx} = A_t \Delta_t A_t^{\mathsf{T}} + V_t$, $D_{xy} = A_t \Delta_t C_t^{\mathsf{T}} = D_{yx}^{\mathsf{T}}$, $D_{yy} = C_t \Delta_t C_t^{\mathsf{T}} + W_t$. Тогда из теоремы 12 следует, что

$$\hat{x}_{t+1} = A\hat{x}_t + A_t \Delta_t C_t^{\mathsf{T}} \left[C_t \Delta_t C_t^{\mathsf{T}} + W_t\right]^{+} (y_t - C\hat{x}_t)\,, \tag{4.3}$$

$$\Delta_{t+1} = A_t \Delta_t A_t^{\mathsf{T}} + V_t - A_t \Delta_t C_t^{\mathsf{T}} \left[C_t \Delta_t C_t^{\mathsf{T}} + W_t\right]^{+} C_t \Delta_t A_t. \tag{4.4}$$

Начальными данными для этой системы является пара $\hat{x}_0 = \bar{x}_0$, Δ_0.

Соотношения (4.3),(4.4), позволяющие рекуррентным образом выписать оптимальную оценку и ее погрешность, известны как уравнения *фильтра Калмана — Бьюси*.

Задача об оптимальном выслеживании зайцев

Для большей наглядности исследуемых в данном примере формальных процедур оптимального оценивания рассмотрим следующую игровую ситуацию.

Охотник выслеживает зайца. Местонахождение зайца есть неизвестная точка x на числовой прямой. Охотник располагает следующей информацией: априорными сведениями и данными наблюдений.

Априорные сведения состоят из $\bar{x}$ — априорной оценки x и $\Delta = \mathbf{E}(\bar{x} - x)^2$ — погрешности этой оценки. Данные наблюдений — последовательность известных величин $y_1, y_2, \ldots, y_t, \ldots$, связанных с неизвестным x соотношениями

$$y_t = x + \xi_t,$$

где ξ_t — ошибки наблюдений — некоррелированные величины с известными статистическими характеристиками:

$$\mathbf{E}(\xi_t) = 0\,,\ \mathbf{E}\left(\xi_t^2\right) = v.$$

Охотник, выслеживая зайца, может строить различные оценки его местонахождения. Если ограничиться одним наблюдением (например, первым) и положить

$$\hat{x} = y_1, \tag{4.5}$$

то погрешность такой оценки равна

$$\mathbf{E}(\hat{x} - x)^2 = \mathbf{E}(\xi_1)^2 = v.$$

При этом игнорируется как априорная информация $(\bar{x}, \Delta)$, так и остальные данные $(y_2, \ldots, y_t, \ldots)$. Рассмотрим три случая.

1. Процедура оптимального оценивания.

Для оптимальной оценки $\hat{x}_t$ неизвестного x по данным $\bar{x}, y_1, \ldots, y_t$ и погрешности $\Delta_t = \mathbf{E}(\hat{x}_t - x)^2$ справедлива [см. (3.1),(3.2)] формула

$$\hat{x}_{t+1} = \hat{x}_t + f_t(y_{t+1} - \hat{x}_t) \quad , \quad \hat{x}_0 = \bar{x}, \tag{4.6}$$

$$\Delta_{t+1} = \Delta_t - \frac{\Delta_t^2}{\Delta_t + v} = \varphi(\Delta_t) \quad , \quad \Delta_0 = \Delta, \tag{4.7}$$

$$f_t = \frac{\Delta_t}{\Delta_t + v} \quad , \quad \varphi(\Delta) = \Delta - \frac{\Delta^2}{\Delta + v} = \frac{\Delta v}{\Delta + v} = v - \frac{v^2}{\Delta + v} \; ,$$

Справедлива явная формула

$$\Delta_t = \frac{v}{t + \dfrac{v}{\Delta_0}}$$

Какой бы ни была погрешность Δ_0 априорной оценки $\bar{x}$, начиная с $t = 1$ и далее, все оценки $\bar{x}_t$ имеют погрешность меньшую v, т.е. будут точнее оценки (4.5). Более того, погрешности Δ_t убывают и стремятся к нулю:

$$\lim_{t \to \infty} \Delta_t = 0.$$

Пусть $\Delta_0 \to \infty$. При этом

$$\Delta_t = \frac{v}{t} \quad , \quad f_t = \frac{1}{t+1},$$

$$\bar{x}_1 = \bar{x}_0 + (y_1 - \bar{x}_0) = y_1,$$

$$\bar{x}_2 = \bar{x}_1 + \frac{1}{2}(y_2 - \bar{x}_1) = \frac{1}{2}(y_1 + y_2),$$

$$\dots$$

$$\bar{x}_t = \frac{1}{t}(y_1 + \ldots + y_t).$$

Как видим, формула (4.6) дает рекуррентное представление оценки, полученной по правилу среднего арифметического от наблюдений.

Прочитав исследование случая 1, заяц, по-видимому, захочет изменить ситуацию в свою пользу и начнет двигаться. Будем считать, что он прыгает в разные стороны случайным образом. При этом его местоположение в последовательные моменты времени удовлетворяет уравнению

$$x_{t+1} = x_t + \eta_t, \tag{4.8}$$

где η_t — величины прыжков — некоррелированные случайные величины со следующими статистическими характеристиками:

$$\mathbf{E}(\eta_t) = 0 \quad , \quad \mathbf{E}(\eta_t^2) = w.$$

В данном случае наблюдения y_t будут связаны с переменными неизвестными значениями x_t соотношениями

$$y_t = x_t + \xi_t.$$

2. Точность построенной процедуры оценивания для движущегося зайца.

Рассмотрим случай, когда заяц движется по закону (4.8), а охотник, не заметив этого (или проигнорировав), продолжает использовать алгоритм (4.6),(4.7), так хорошо зарекомендовавший себя в случае 1. Разумеется, в этих новых условиях фильтр (4.6) будет формировать оценки $\bar{x}_t$, отличные от $\bar{x}_t$.

Из равенств

$$\bar{x}_{t+1} - x_{t+1} = \bar{x}_t + f_t\,(y_{t+1} - x_t) - x_{t+1} =$$

$$= \bar{x}_t + f_t\,(x_t + \eta_t + \xi_{t+1} - x_t) - x_t - \eta_t =$$

$$= (1 - f_t)(\bar{x}_t - x_t) + (f_t - 1)\eta_t + f_t\xi_{t+1}$$

следуют рекуррентные уравнения

$$\Delta_{t+1} = (1 - f_t)^2(\Delta_t + w) + f_t^2 v \tag{4.9}$$

для погрешности $\Delta_t = \mathbf{E}(x_t - x_t)^2$.

Положив в (4.9) $w = 0$, получим уравнения

$$\Delta_{t+1} = (1 - f_t)^2\Delta_t + f_t^2 v \tag{4.10}$$

для погрешности Δ_t случая 1.

Из (4.9),(4.10) следуют уравнения

$$r_{t+1} = (1 - f_t)^2(r_t + w)$$

для $r_t = \Delta_t - \Delta_t$.

Замена $\delta_t = \frac{1}{w}(r_t + w)$ приводит к уравнению

$$\delta_{t+1} = (1 - f_t)^2\delta_t + 1$$

Заметив, что

$$(1 - f_t)^2 = \frac{t^2}{(t+1)^2},$$

и сделав замену $\beta_t = t^2\delta$, получим

$$\beta_{t+1} = \beta_t + (t+1)^2. \tag{4.11}$$

Из (4.11) находим

$$\beta_t = 1^2 + 2^2 + \ldots + t^2 = \frac{t(t+1)(2t+1)}{6}$$

В результате получаем

$$\Delta_t = \Delta_t + r_t,$$

где $\delta_t \to 0$, а

$$r_t = w\left(\frac{(t+1)(2t+1)}{6t} - 1\right) \to \infty.$$

Таким образом, при любом сколь угодно малом w процедура, оптимальная в случае 1, для случая 2 будет расходиться.

Прочитав исследования случая 2, охотник, по-видимому, захочет изменить ситуацию в свою пользу и постарается учесть в процедуре оценивания динамику зайца.

3. Оптимальное оценивание местонахождения движущегося зайца.

Для оптимальной оценки $\tilde{x}_t$ неизвестного x_t с учетом динамики (4.8) справедливы формулы (фильтр Калмана – Бьюси)

$$x_{t+1} = x_t + \frac{\tilde{\Delta}_t + w}{\tilde{\Delta}_t + w + v}\left(y_{t+1} - \tilde{x}_t\right) \quad , \quad \tilde{x}_0 = x,$$

$$\tilde{\Delta}_{t+1} = \tilde{\Delta}_t + w - \frac{\left(\tilde{\Delta}_t + w\right)^2}{\tilde{\Delta}_t + w + v} = \varphi\left(\tilde{\Delta}_t + w\right) \quad , \quad \tilde{\Delta}_0 = \Delta_0.$$

Какой бы ни была погрешность Δ_0, последовательность $\tilde{\Delta}_t$ сходится к стационарному значению

$$\Delta^* = v\frac{2}{1 + \sqrt{1 + u\dfrac{v}{w}}}$$

– решению уравнения $\Delta = \varphi(\Delta + w)$.

Максимум, чего может добиться заяц, увеличивая w – дисперсию прыжков, это сделать $\Delta^* = v$. При этом возможности охотника минимизировать Δ^* связаны с необходимостью уменьшать v – дисперсию ошибок наблюдений.

Список литературы

[1] Воеводин В. В. Вычислительные основы линейной алгебры. М.; Наука, 1977.

[2] Гантмахер Ф. Р. Теория матриц. М.; Наука, 1967.

[3] Воеводин В. В., Кузнецов Ю. А. Матрицы и вычисления. М.; Наука, 1984.

[4] Алберт А. Регрессия, псевдоинверсия и рекуррентное оценивание. М.; Наука, 1977.

[5] Рао С. Р. Линейные статистические методы и их применение. М.; Наука, 1968.

[6] Себер Дж. Линейный регрессионный анализ. М.; Мир, 1980.

[7] Лоусон Ч., Хенсон Р. Численное решение задач метода наименьших квадратов. М.; Наука, 1986.

[8] Линник Ю. В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. М.; Физматгиз, 1962.

[9] Липцер Р. Ш., Ширяев А. Н. Статистика случайных процессов. М.; Наука, 1974.

[10] Балакришнан А. В. Теория фильтрации Калмана. М.; Мир, 1988.

[11] Браммер К., Зиффлинг Г. Фильтр Калмана-Бьюси. М.; Наука, 1982.

[12] Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. М.; Наука, 1979.

[13] Иванов В. К., Васин В. В., Танана В. П. Теория линейных некорректных задач и ее приложения. М.; Наука, 1978.

Оглавление

Лев Борисович Ряшко

Оптимальное рекуррентное оценивание

Учебное пособие

Редактор **Т.А. Сасина**
Технический редактор **Э.А. Максимова**

ЛР N 020257 от 22.11.96

Подписано в печать 16.06.97. Формат 60 × 84 1/16.
Бумага для множительных аппаратов. Печать офсетная.
Уч.-изд.л. 3,0. Усл.печ.л. 3,0. Заказ 265. Тираж 500 экз.
Уральский государственный университет
им. А.М. Горького.
Екатеринбург, пр. Ленина, 51

Типолаборатория УрГУ. Екатеринбург, пр. Ленина, 51